frontiers in Genetics

REVIEW
published: 28 October 2020
doi: 10.3389/fgene.2020.591194

# Mapping Regulatory Determinants in Plants

*Mary Galli[1]*, Fan Feng[1] and Andrea Gallavotti[1,2†]*

*[1]Waksman Institute of Microbiology, Rutgers University, Piscataway, NJ, United States, [2]Department of Plant Biology, Rutgers University, New Brunswick, NJ, United States*

The domestication and improvement of many plant species have frequently involved modulation of transcriptional outputs and continue to offer much promise for targeted trait engineering. The cis-regulatory elements (CREs) controlling these trait-associated transcriptional variants however reside within non-coding regions that are currently poorly annotated in most plant species. This is particularly true in large crop genomes where regulatory regions constitute only a small fraction of the total genomic space. Furthermore, relatively little is known about how CREs function to modulate transcription in plants. Therefore understanding where regulatory regions are located within a genome, what genes they control, and how they are structured are important factors that could be used to guide both traditional and synthetic plant breeding efforts. Here, we describe classic examples of regulatory instances as well as recent advances in plant regulatory genomics. We highlight valuable molecular tools that are enabling large-scale identification of CREs and offering unprecedented insight into how genes are regulated in diverse plant species. We focus on chromatin environment, transcription factor (TF) binding, the role of transposable elements, and the association between regulatory regions and target genes.

**Keywords: plant genomics, transcriptional regulation, chromatin, transcription factor binding, cis-regulatory regions**

OPEN ACCESS

***Edited by:***
*Wenqin Wang,*
*Shanghai Jiao Tong University, China*

***Reviewed by:***
*Qi Li,*
*Center for Excellence in Molecular Plant Sciences (CAS), China*
*Yubing He,*
*Nanjing Agricultural University, China*

****Correspondence:***
*Mary Galli*
*marygalli@waksman.rutgers.edu*

***†ORCID:***
*Andrea Gallavotti*
*orcid.org/0000-0002-1901-2971*

***Specialty section:***
*This article was submitted to Evolutionary and Population Genetics, a section of the journal Frontiers in Genetics*

***Received:*** *04 August 2020*
***Accepted:*** *28 September 2020*
***Published:*** *28 October 2020*

***Citation:***
*Galli M, Feng F and Gallavotti A (2020) Mapping Regulatory Determinants in Plants. Front. Genet. 11:591194. doi: 10.3389/fgene.2020.591194*

## REGULATORY REGIONS AND MECHANISMS REVEALED BY CLASSIC STUDIES

Mining trait-associated genetic factors has traditionally been performed using classical genetics, GWAS, and QTL analysis. Examples from these studies serve as excellent guides for understanding the molecular basis of phenotypic diversity (Deplancke et al., 2016). In particular, the regions corresponding to several beneficial traits associated with the domestication and diversification of many plant species from their wild relatives have been mapped by these approaches and frequently shown to be located in the intergenic space, sometimes residing up to 100 kb from the closest protein coding genes (**Figure 1A**; Olsen and Wendel, 2013; Rodgers-Melnick et al., 2016; Swinnen et al., 2016; Lu et al., 2019). Correspondingly, these traits involve variations in gene expression, with variants affecting either the level of expression or the spatial and/or temporal pattern of expression of certain genes (**Figure 1B**; Meyer and Purugganan, 2013; Springer et al., 2019). Unlike changes to protein-coding genes which often result in easily interpretable loss-of-function alleles, the exact causative features underlying functional cis-regulatory

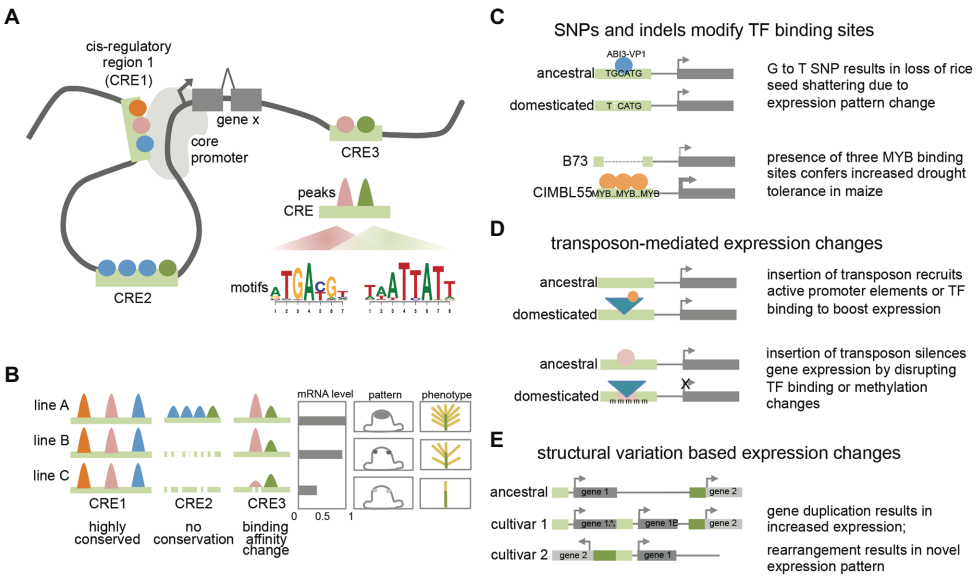


**FIGURE 1 |** Plant transcriptional regulation **(A)** model of plant transcriptional regulation at gene X. Colored circles represent different TFs binding to three distinct cis-regulatory regions (CREs; light green bars) that can contact the core promoter *via* DNA looping. Motifs enriched within binding peaks for two TFs are shown for CRE3. **(B)** Conservation and variation of TF binding events among different lines or accessions. Colored peaks represent different TF binding events within CREs. mRNA expression levels, cell-type specific expression pattern, and resulting phenotype are shown. **(C)** Examples showing how single nucleotide polymorphisms (SNPs) and indels can result in expression and phenotypic changes. **(D)** Examples showing how transposon insertions can result in expression and phenotypic changes. **(E)** Examples showing how structural variants can result in expression changes.

regions (CREs) are currently difficult to identify given the variable nature of regulatory elements, their frequent gene-distal location, and the lack of an obvious rigid code that determines their functionality. Understanding the molecular nature of these changes however lies at the heart of our ability to accelerate crop improvement using CRISPR-based targeted engineering of useful traits and traditional breeding (Rodríguez-Leal et al., 2017; Chen et al., 2019; Eshed and Lippman, 2019; Springer et al., 2019).

In several cases, the molecular nature of the phenotypic variation has been determined and found to be associated with a range of different causes. These include single nucleotide polymorphisms (SNPs) that affect transcription factor (TF) binding, either by disrupting or recruiting additional TF binding sites. For example, a G to T nucleotide change located 12 kb upstream of the *qSH1* gene in rice, a BEL-type homeobox TF, is believed to disrupt an ABI3-VP1 TF binding site (Konishi et al., 2006). This results in a loss of *qSH1* expression in the pedicel abscission zone and a subsequent non-shattering phenotype that facilitated higher harvesting yields. Alternatively, changes in TF binding can also involve advantageous gain of function elements. A GWAS screen for drought tolerance in maize identified a 366 bp region located in the proximal upstream region of *ZmVPP1*, a vacuolar-type $H^+$-pyrophosphatase, that conferred increased drought tolerance in several varieties (Wang et al., 2016). This fragment contains three putative MYB binding sites, which were shown to increase expression of *ZmVPP1* relative to the drought-sensitive maize line B73, which lacks the MYB binding sites.

In other cases, functional traits associated with cis-regulatory elements (CREs) may not involve nucleotide variations that directly correspond to known TF binding sites but are instead located nearby. This is the case for the rice *GW7* gene, which affects grain width and grain quality (Wang et al., 2015b). Certain rice varieties were found to contain two short indels directly adjacent an SBP16/GW8 TF binding motif in the proximal upstream region of *GW7*. These indels do not directly disrupt the TF binding motif but do appear to lower expression of *GW7* relative to varieties in which the indels are absent. Given that regulatory regions typically contain multiple different TF binding sites (Hardison and Taylor, 2012; Ricci et al., 2019), such examples could indicate that these divergent regions simply correspond to unknown TF binding sites and reflect the incompleteness of TF binding motif characterization in plants. Alternatively, they could alter local DNA shape (i.e., the sequence-dependent DNA structure surrounding the motif) or spacing between adjacent motifs, among other factors that contribute to the complexity of TF binding specificity (Slattery et al., 2014). Such examples highlight the need for comprehensive annotation of TFs and other regulatory regions. Similar examples have been noted in non-plant studies, where there is accumulating evidence that causative SNPs frequently do not directly affect TF binding motifs, but may impact cooperative or collaborative binding of TF complexes (Deplancke et al., 2016).

Transposon insertions in regulatory regions can also influence gene expression of adjacent genes, resulting in either elevated or suppressed gene expression levels, and likely act through a variety of mechanisms (Hirsch and Springer, 2017; Zhao et al., 2018).

A classic example of the former in plants is the presence of a *Hopscotch* element located ~60 kb upstream of the *TEOSINTE BRANCHED1 (TB1)* gene, a TCP-family TF that determines the apical dominance of domesticated maize relative to its highly branched wild ancestor teosinte (Studer et al., 2011). The *Hopscotch* element enhances the expression of *TB1* through an unknown mechanism. Interestingly, a nearby *Tourist* transposon within the same enhancer appears to repress expression of *TB1*, highlighting the dynamic nature of transcriptional changes conferred by transposable elements. Another illustrative example includes the insertion of a *Copia* retroelement in the proximal upstream region of the *RUBY* gene in blood oranges. *RUBY* encodes a MYB TF involved in anthocyanin production and its expression level is increased by cold-induced expression conferred by sequences within the long terminal repeat (LTR) that are hypothesized to harbor either promoter-like features with a TATA box and TSS, or other upstream activating sequences (Butelli et al., 2012). These examples suggest that like other cases from animals, transposons may act as novel promoters by recruiting the basal transcriptional machinery or introducing tissue-specific TF binding sites (or disrupting repressive TF binding sites; Butelli et al., 2012; Sundaram et al., 2014).

Transposon insertions within regulatory regions are also able to negatively impact gene expression. They can do this by disrupting existing TF binding sites or other regulatory features, or *via* epigenetic changes typically involving repressive DNA methylation (Huang and Ecker, 2018). For example, one of the major factors determining fruit color in grape species, is caused by a *Gypsy-like* retrotransposon insertion, *Gret1*, in the upstream region of *MYBA1*, involved in berry anthocyanin production. As opposed to the *RUBY* blood orange case described earlier, the presence of *Gret1* results in loss of gene expression and the white-colored berries typical of chardonnay (Kobayashi et al., 2004). Similar cases of transposon mediated gene repression are also seen in maize at the *ZmCCT10* and *ZmCCT9* loci, two genes involved in flowering-time regulation whose causative transposon insertions reside 2.5 and 57 kb upstream, respectively (Yang et al., 2013; Huang et al., 2017). In general, the mechanisms of how such transposon associated CREs influence expression are not fully understood although these examples and others suggest they can affect both distal enhancers and proximal regulatory regions. In other cases involving transposon insertions in regulatory regions, changes in DNA methylation have been documented as the underlying cause of stable gene downregulation (Hirsch and Springer, 2017). Examples of such epialleles include a methylated *hAT* element inserted in the proximal regulatory region of the melon *CmWIP* gene, which controls sex determination (Martin et al., 2009) and a SINE retrotransposon inserted upstream of the tomato *VTE3* gene, involved in vitamin E biosynthesis (Rossi et al., 2014). Possible mechanisms that explain stable transposon-triggered repression include spreading of methylation marks from the TE into the adjacent regulatory region, thus altering chromatin accessibility or blocking TF motif binding (many TFs preferentially bind unmethylated sites; Eichten et al., 2012; O'Malley et al., 2016; Huang et al., 2018). Overall, these examples as well as studies analyzing global transposon location (i.e., 86% of maize genes contain a TE within 1 kb of the gene; Hirsch and Springer, 2017) and association with eQTL, suggest that TE-driven transcriptional influence is frequent and in certain genomes may be major drivers of regulatory variation (Zhao et al., 2018; Noshay et al., 2020).

Although far less frequent than regulatory changes associated with TE insertions, there are several reports of regulatory epialleles that appear to have formed spontaneously. These include the *Colorless non-ripening (Cnr)* mutant allele of tomato, which encodes an SBP TF that affects color ripening (Manning et al., 2006). In the *Cnr* mutant, the upstream regulatory region of the *Cnr* gene is stably hypermethylated throughout development, leading to reduced expression of the gene (Zhong et al., 2013). Interestingly, the methylated sites are adjacent to two MADS-box TF binding sites bound by RIPENING INHIBITOR1 (RIN1; a MADS-box TF) in ChIP-seq (Zhong et al., 2013) suggesting that methylation changes in the *Cnr* epimutant could impact TF binding.

Finally, structural variants have also been shown to affect regulatory outputs by altering gene copy number and/or the arrangement or composition of CREs (Alonge et al., 2020), highlighting the modular architecture of regulatory elements. In the case of inversions, a certain gene may become located adjacent to an otherwise distally located gene or regulatory region and assume novel expression patterns. This appears to be the case for the classic *Tunicate* allele of maize, which shows unusually long glumes in both inflorescences as a result of ectopic expression from the 3' region of a gene normally located 1.8 Mb away (Han et al., 2012). Other structural variants include segmental duplications that increase gene copy number. While these do not directly involve changes in CREs, they do appear to be a subtle but possibly frequent mechanism of trait-associated transcriptional modulation in certain species (Alonge et al., 2020). Other situations in which putative regulatory regions are rearranged or duplicated are less clear. A good example of this is the ~4 kb DICE distal enhancer element in maize which confers increased expression of the *BX1* gene and consequently increased herbivore resistance (Betsiashvili et al., 2015; Zheng et al., 2015). The DICE element appears to be a divergent duplication of nearby sequences, and the increased expression may result from increased recruitment of specific TFs (Galli et al., 2018). Additional examples from maize include the classic cases of the *b1* and *Vgt1* loci, both of which are associated with structural variation in distal non-coding regions that results in epigenetic changes (Stam et al., 2002; Castelletti et al., 2014).

Detailed genetic and molecular characterization of QTL and classic cases have established a solid groundwork for understanding how regulatory changes influence many phenotypic traits in plants. However, they likely represent only a small fraction of the genetic variation and molecular mechanisms that govern transcriptional response for quantitative traits. Recently-developed genomics based techniques are paving the way for large-scale mining of putative CREs and begin to outline certain molecular signatures that correlate with gene expression and are conserved across species and accessions

(Maher et al., 2018; Lu et al., 2019; Alonge et al., 2020). Ultimately, combining both genetic and genome-wide studies will prove a powerful technique to better understand beneficial traits.

## GENOME-WIDE IDENTIFICATION OF cis-REGULATORY REGIONS

Regulatory DNA in eukaryotes is generally characterized by chromatin accessibility, low DNA methylation, and is often associated with distinct histone modifications (Marand et al., 2017; Oka et al., 2017; Klemm et al., 2019; Lu et al., 2019). In plants, several recent studies have taken advantage of these properties to mine candidate regulatory elements at the genomic level (Sullivan et al., 2014; Rodgers-Melnick et al., 2016; Oka et al., 2017; Lü et al., 2018; Maher et al., 2018; Lu et al., 2019; Ricci et al., 2019; Parvathaneni et al., 2020). Such approaches are critical because while previous promoter and QTL studies suggest that most regulatory elements appear to lie within 1–2 kb upstream of the gene body in smaller genomes such as *Arabidopsis*, in larger genomes, regulatory regions reside within a much broader upstream area, with distal elements occasionally located hundreds of kb from the genes they regulate, making their identification by traditional means arduous. Therefore, the identification of accessible chromatin regions (ACRs) using techniques such as ATAC-seq (Assay for Transposase-Accessible Chromatin using sequencing), MNaseHS (micrococcal nuclease hypersensitivity), and DNaseHS (DNAse hypersensitivity) has been highly informative for mapping regulatory regions in plants, revealing their frequency, size, and location, as well as many other important aspects. These studies demonstrate that ACRs are most often found near transcription start and end sites, but can also frequently be found over 2–200 kb from any gene depending on the species (Sullivan et al., 2014; Rodgers-Melnick et al., 2016; Oka et al., 2017; Maher et al., 2018; Lu et al., 2019; Ricci et al., 2019). They also show that ACRs can be condition and tissue-specific, highlighting the dynamic nature of chromatin (Sullivan et al., 2014; Rodgers-Melnick et al., 2016; Oka et al., 2017; Maher et al., 2018; Ricci et al., 2019; Parvathaneni et al., 2020). In support of their functionality, most identified ACRs are enriched for TF binding events and motifs and show transcriptional enhancer activity (see below for more detail; Sullivan et al., 2014; Ricci et al., 2019). Importantly, it was shown that SNPs in ACRs explain up to 40% of the variability in quantitative traits in maize and in particular overlap with several classically defined distal QTL discussed previously, substantiating their functionality and highlighting the role of regulatory regions in modulating phenotypes (Rodgers-Melnick et al., 2016; Ricci et al., 2019).

A landmark, cross-species comparative study of 13 angiosperm species with genome sizes ranging from ~100 to 5,000 Mb demonstrated that ACRs account for 0.2–6.5% of the total genome of a species and that their location varies according to genome size (Lu et al., 2019). For example, while the total sequence length of ACRs was fairly consistent across species regardless of genome size, large genomes showed a greater percentage of distally located ACRs (i.e., small genomes such as *Arabidopsis* showed that only ~6% of all ACRs were distal compared to ~46% in barley). Transposon insertions were found to be one of the main factors contributing to this occurrence, presumably pushing ACRs away from genes (Lu et al., 2019). Transposons themselves also appeared to be responsible for creating certain species-specific distal ACRs, as noted previously from classical studies (see above i.e., maize *TB1*). The controlled parallel nature of the Lu et al. (2019) study also allowed several important cross-species observations such as the finding that the number of ACRs correlated with the number of genes within a species and that many distal ACRs were conserved between sister species. Overall, an important finding from this study is that large and small plant genomes appear to be structured differently, despite harboring many of the same genetic pathways and gene regulatory networks (Lu et al., 2019). This underscores the importance of empirically mining sufficient amounts of regulatory information both for direct application in a species of interest such that ultimately such information will enable accurate machine learning predictions in other crop species.

A major factor in the characterization of putative regulatory regions is determining their functionality. In animals, regulatory regions are generally categorized into classes such enhancers, insulators, or promoters depending on their role in gene expression (Andersson and Sandelin, 2020). These terms however remain somewhat ambiguous despite an enormous effort toward their classification, perhaps because the elements themselves are heterogeneous (Andersson and Sandelin, 2020; Gasperini et al., 2020). In plants, these operational definitions are even more vague; however, studies have begun to tease out some common trends. Plant adapted versions of massively parallel promoter and enhancer reporter assays such as self-transcribing active regulatory region sequencing (STARR-seq; Ricci et al., 2019; Jores et al., 2020), show that many ACRs are capable of enhancing gene expression (Ricci et al., 2019). Traditional STARR-seq works by inserting fragments either from randomly sheared genomic sequence, BAC libraries, or small fragments such as those from ATAC-seq and placing them downstream of a cassette containing a minimal promoter fused to GFP (Arnold et al., 2013). Because enhancers are assumed to be capable of controlling gene expression regardless of distance or orientation (according to the classical definition), STARR-seq allows for self-driven transcription of the element and quantitative readout. In maize, both proximal and distal ACRs were found to show a general enhancement of activity, relative to randomly selected regions with similar features (Ricci et al., 2019). On the other hand, a modified version of STARR-seq using transient transfection in tobacco leaves found that four known plant enhancers gave the strongest transcriptional output when placed immediately upstream of a minimal promoter and were not active when placed in the 3'UTR of the reporter gene (Jores et al., 2020). Further studies are needed to tease out the functional determinants and optimal architecture of the various classes of regulatory elements. Given their utility to generate

synthetic transcriptional units for agricultural improvement (Liu and Stewart, 2016), findings from such assays, and approaches could be directly applicable in plants, unlike in animals.

Genomes also typically harbor specific chromatin features that serve as another potential source of regulatory information (Marand et al., 2017). In animals, ACRs are often associated with distinct histone modifications that correlate with gene expression outputs (Hardison and Taylor, 2012; Gasperini et al., 2020). There has been much focus placed on using unique signatures of these various chromatin marks to identify particular classes of regulatory elements (e.g., enhancers) to aid genome annotation efforts and understand how chromatin environment impacts gene expression. However, it is widely accepted that operational definitions based on these biochemical marks serve as a guide rather than a fixed rule (Gasperini et al., 2020). Several large-scale studies have profiled histone modifications in various plant species (Oka et al., 2017; Lü et al., 2018; Lu et al., 2019; Peng et al., 2019; Ricci et al., 2019), and detailed analysis suggests that as in animals, certain chromatin signatures correlate with gene expression levels: expressed genes are enriched for H3K4me3, H3K56ac, and H2A.Z at the transcription start site, whereas repressed genes are enriched for H3K27me3 and H2A.Z (Lu et al., 2019). Furthermore, in maize, it appears that H3K27me3 marks often correspond to tissue-specific genes while H3K4me1 and H3K4me3 tend to mark broadly expressed genes (Lu et al., 2019; Peng et al., 2019; Ricci et al., 2019). Combining histone modification data with ACRs found that H3K9/K27/K56ac marks were generally associated with high expression levels of nearby genes and may represent enhancers. Distal ACRs instead marked by H3K27me3 tended to be located near genes with lower levels of expression and may represent repressor elements. Interestingly, it appears that some plant histone modification trends differ from those found in animals (Lu et al., 2019). For example, while H3K4me1 marks are typically found at distal CREs, in plants, this modification was not frequently associated with distal CREs (Lu et al., 2019).

Finally, DNA methylation maps are also highly valuable for mining regulatory information (Crisp et al., 2019). Prior studies have noted that most ACRs are hypomethylated, and in large genomes that are typically heavily methylated, unmethylated regions (UMRs) serve as an excellent tool to mine functional CREs (Crisp et al., 2020). Importantly, UMRs tend be static across most tissues and conditions in plants, whereas ACRs and histone modifications are often dynamic. Therefore, UMRs from a single tissue can be used to locate CREs, and when paired with chromatin accessibility data from a dissimilar tissue, can reveal CREs potentially set to become accessible or expressed in another tissue (Crisp et al., 2020).

Overall, these various genome-wide approaches for mining regulatory elements are generating highly informative maps that are crucial for understanding regulatory dynamics (**Figure 2**). Such data are critical for locating regulatory regions for use in transgenic studies or harnessing tissue-specific promoters for genetic engineering purposes.

## TRANSCRIPTION FACTORS: DRIVERS OF GENE EXPRESSION

At the heart of transcriptional regulation is DNA-binding TFs and TF complexes bound to CREs. Transcription factors recognize short DNA sequence motifs in regulatory regions of their target genes and control the gene expression changes responsible for plant developmental programs and environmental responses. TFs bind to family-specific DNA motifs that contain four to six nucleotides, although many instances of longer and more

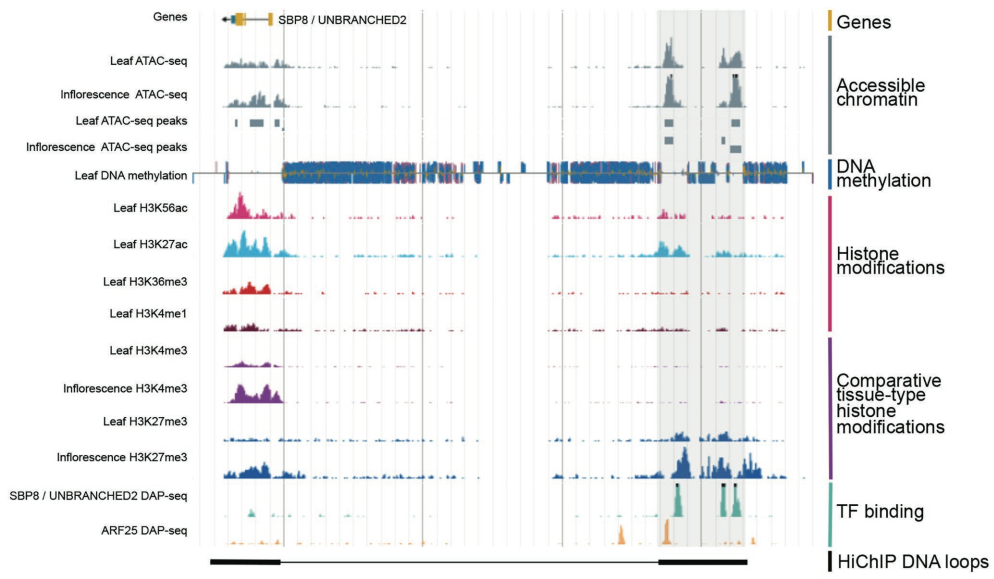


**FIGURE 2 |** Integration of various types of genomic regulatory data allows for the identification of CREs. Shown is a genome browser view of putative distal CRE (gray shaded region) located 40 kb upstream of the SBP8/UNBRANCHED2 gene in maize. Data obtained from Ricci et al., 2019.

complex architecture are known (Jolma et al., 2013; Weirauch et al., 2014; O'Malley et al., 2016). Particularly, in the case of short motifs, it is clear that TFs do not bind to all instances of these motifs within a given genome, suggesting that other factors also influence binding specificity (Hardison and Taylor, 2012; Todeschini et al., 2014). These have been shown to include DNA shape, i.e., the DNA sequence surrounding the motif, which is not directly bound by the TF (Slattery et al., 2014), as well as other factors such as the presence of proximally located motifs that can be bound by cooperating TFs (Deplancke et al., 2016). However, while these features play a role, the precise determinants of TF binding specificity remain unclear. One of the many additional interesting features of TF binding is the tendency for diverse TFs to bind in clusters, often lying within a region of open chromatin (**Figure 1**; Gasperini et al., 2020). This has been observed in many animal systems where a large number of genome-wide TF binding maps are available, and appears to occur in plants as well (see below for more detail). It remains unclear how these clusters of TFs are involved in gene regulation; however, the modular/combinatorial binding nature of these regulatory regions (i.e., multiple TFs binding) appears to allow genes to be controlled in tissue-specific or temporal manner (Spitz and Furlong, 2012). In plants, this is particularly intriguing from an agronomic engineering perspective because it suggests that phenotypes associated with distinct organs (i.e., ear traits but not tassel traits in maize) could be separated, allowing specific alterations to one organ or conditional response without altering another with a less desirable phenotype (Dong et al., 2019).

There are several methods by which to identify TF binding. ChIP-seq is the current gold-standard method for determining *in vivo* binding sites of TFs in live cells (Johnson et al., 2007). This method enables the identification of genomic binding sites in a tissue-specific chromatin context with high resolution (Park, 2009; Kaufmann et al., 2010). DNA-protein complexes are immunoprecipitated using an antibody specific to the protein of interest or a tag that is fused to the protein, and DNA is purified from the immunoprecipitated complex and subjected to next-generation sequencing. Several key factors that contribute to high-quality data in ChIP-seq, include antibody selection, negative controls, and biological replicates (Park, 2009; Kidder et al., 2011; Landt et al., 2012). Because of its *in vivo* context, ChIP-seq captures DNA bound both directly and indirectly by the TF of interest. This can include sites bound by hetero- or multimeric complexes. Many small and medium scale ChIP-seq studies have been carried out in *Arabidopsis* in contrast to the handful that have been performed in larger genomes such maize and soybean (Bolduc et al., 2012; Huang et al., 2012; Gregis et al., 2013; Lau et al., 2014; Tsuda et al., 2014; Li et al., 2015; Pautler et al., 2015; Jung et al., 2016; Song et al., 2016; Feng et al., 2018; Jo et al., 2020). A major limitation to ChIP-seq in plants is the time and effort required to either create transgenic lines or generate antibodies.

Performing ChIP-seq using protoplasts that transiently express epitope-tagged transcription factors is an alternative approach (Kong et al., 2012; Lee et al., 2017; Tu et al., 2020), as in some cases, specific antibodies against an endogenous protein of interest or transgenic lines expressing the protein of interest fused with a tag in a mutant background are unavailable. Protoplasts can be obtained either from mesophyll or other tissues such as root or stem and are transformed with a plasmid that expresses the protein of interest fused with an epitope-tag driven by a ubiquitously expression promoter such as 35S (Hernandez et al., 2007; Yoo et al., 2007; Kong et al., 2012; Para et al., 2014). ChIP-seq using protoplasts has obvious advantages as it bypasses the requirements for antibody or transgenic plants; however, overexpression of proteins in protoplasts might lead to altered genomic binding profiles due to excess protein in the cell (Kidder et al., 2011). A recent large-scale study using this approach in maize to map the binding sites of 104 TFs in leaves observed several key findings. As seen in animals, plant TF binding sites clustered together, covering ~2% of the maize genome and reinforcing the emerging paradigm that multiple TFs are needed for regulation of a single locus (Tu et al., 2020). These results also suggest co-binding appears to be important for TF specificity in maize (Tu et al., 2020).

Another modified version of ChIP-seq is cleavage under targets and release using nuclease (CUT&RUN), a chromatin profiling strategy in which antibody-targeted controlled cleavage by micrococcal nuclease releases specific protein-DNA complexes into the supernatant for paired-end DNA sequencing (Skene and Henikoff, 2017; Skene et al., 2018). Compared to ChIP-seq, CUT&RUN has several key advantages such as no crosslinking, which avoids false positive signals; *in situ* targeted digestion, which greatly reduces background; efficiency, as it can be finished in a day; and high signal-to-noise ratio, requiring only one tenth of the sequencing depth as ChIP-seq.

DAP-seq is an *in vitro* alternative to ChIP-seq (O'Malley et al., 2016). DAP-seq works by combining a standard Illumina-based genomic DNA sequencing library together with an *in vitro* expressed affinity-tagged TF coupled to magnetic beads. After a series of washes, TF-bound DNA is eluted, enriched, and barcoded for multiplexing, followed by next-gen sequencing (Bartlett et al., 2017). Resulting reads produce genome-wide peak maps similar to ChIP-seq, but often with higher resolution. A main advantage of DAP-seq is that it combines the low cost and high throughput of an *in vitro* assay with DNA in its native sequence context thereby preserving DNA structure and DNA methylation marks that are known to impact TF binding (O'Malley et al., 2016). Bound fragments are directly mapped to a genome unlike other *in vitro* assays such as HT-SELEX and protein binding microarrays, which report only motifs (Jolma et al., 2013; Weirauch et al., 2014). DAP-seq has been used to generate high quality peak maps for 529 *Arabidopsis* TFs and several maize TFs (O'Malley et al., 2016; Galli et al., 2018; Ricci et al., 2019). This data revealed many informative properties of plant TFs such as high frequency at which TFs from the same family- or subfamily-type bind similar sites, that TFs bind a very small fraction of all motif instances, and again that TFs cluster together in proximal promoters (and distal enhancers which are often located over 20–100 kb from their putative target gene in maize). Comparative studies of DAP-seq showed significant overlap with ChIP-seq data;

however, DAP-seq generally produces more peaks than ChIP-seq suggesting that DAP-seq captures binding events that take place independent of tissue- or condition-specific chromatin information (O'Malley et al., 2016).

Genome wide TF binding maps generated by these various techniques will be essential for understanding factors influencing both TF binding and TF activity. Yet while TFs are the major modulators of transcriptional activity, and their individual importance is often evident from mutations with severe developmental consequences, how TFs actually modulate gene expression remains largely unclear (de Boer et al., 2020). As in animal systems, it is also clear that not all TF binding is functional (Spitz and Furlong, 2012; Para et al., 2014; Brooks et al., 2019; Gasperini et al., 2020). Therefore, another challenge will be establishing determinants of TF activity and how these are influenced by factors such as position of binding sites, binding site strand, helical position, and protein interactions (de Boer et al., 2020). As mentioned previously, TF binding sites often cluster together and form cis-regulatory modules (CRMs; Hardison and Taylor, 2012) which themselves could impact TF activity. These CRMs and the individual TF binding sites within are often conserved within and across species indicating that together they may be important for TF activity and gene expression. Deciphering the degree to which plant TFs may work cooperatively will require dissection of CRMs using both natural variation and targeted genomic editing to better understand these regulatory regions.

## INTERACTIONS BETWEEN REGULATORY REGIONS AND GENES: TARGET GENE IDENTIFICATION AND FUNCTIONAL CONSEQUENCES OF 3D CONFORMATION

An essential aspect of mining regulatory elements in any genome is being able to associate a putative regulatory region with a target gene or genes, and its expression dynamics. This remains a particularly challenging task in large genomes where regulatory regions may be located hundreds of kb away (Pliner et al., 2018). The current model of regulatory region-gene interactions involves looping of DNA in 3D space to allow physically distant regions to contact core promoters (**Figure 1A**; Shlyueva et al., 2014), and until recently this general eukaryotic model was derived largely from data in animals. Several plant studies using chromosome conformation capture (3C)-based techniques such as Hi-C and other variants, which capture global chromatin interactions (van Steensel and Dekker, 2010), have now shown that plant 3D chromatin organization generally resembles that reported in animals (Wang et al., 2015a, 2017; Dong et al., 2017; Liu et al., 2017; Mascher et al., 2017; Li et al., 2019; Peng et al., 2019; Ricci et al., 2019; Sun et al., 2020), despite the absence of certain proteins such as CTCF that are associated with this phenomenon in animals (Liu et al., 2017; Rowley et al., 2017). In these assays, chromatin contacts within a particular tissue are first cross-linked with formaldehyde, sheared to linearize the DNA, and then DNA ends are ligated together. The resulting ligated DNA is sequenced and consists of fragments that may not reside close in linear genomic space but are contacted in 3D space, often reflecting long-range spatial associations. Importantly, comparison among various plant genomes suggests that the 3D architecture of small, compact plant genomes such as *Arabidopsis* which tend to have CREs located within or near genes, differs from that of larger plant genomes which often form extensive long-range chromatin loops (Wang et al., 2015a, 2017; Dong et al., 2017; Liu et al., 2017; Ricci et al., 2019).

Bulk chromatin capture techniques such as Hi-C are often limited in their resolution, preventing the detailed empirical mapping of linkages between regulatory regions and target genes, and thus limiting the functional mapping of regulatory elements. More focused techniques such as Hi-ChIP and ChIA-PET use antibodies to enrich for a specific subset of chromatin interactions that are associated with RNA polymerase II, a particular histone modification, or transcription factor, offering greater resolution at a lower sequencing depth (Fullwood et al., 2009; Mumbach et al., 2016). A series of reports that mapped 3D chromatin interactions using several different higher-resolution assays in maize, a model species that is likely representative of many large crop genomes, revealed the importance of chromatin loops for influencing gene expression and phenotype (Li et al., 2019; Peng et al., 2019; Ricci et al., 2019; Sun et al., 2020). Collectively, these studies indicated that: (i) interactions between genes and proximal (<2 kb) and distal (>20 kb) ACRs (i.e., putative CREs) were common, and confirmed many genetically identified long-distance regulatory regions; (ii) genes with chromatin interactions associated with active promoters and enhancers tended to have higher expression levels than those without; (iii) functional CRE-gene interactions showed a strong loop signal intensity and tended to lie directly upstream of the gene (i.e., gene skipping was less common than direct contact; Ricci et al., 2019); (iv) gene pairs connected by loops within their proximal promoters were often transcriptionally coordinated; (v) tissue-specific (i.e., ear vs. shoot) proximal-distal interactions correlated with tissue-specific gene expression; and (vi) genes and CREs were often connected by multiple loops suggesting a complex pattern of regulation. Many of these features are likely to be conserved in other plant genomes and serve as a foundation for predicting functional regulatory elements in other species. However, given the vast diversity and size differences among plant genomes, and the prevalence of polyploidy among domesticated crop species, it is possible that many species exhibit unique chromatin conformation features that influence gene expression and certain species-specific traits (Wang et al., 2017; Concia et al., 2020).

Overall, these studies in plants confirm that long-range contacts do frequently occur in plants and raise many additional intriguing aspects of gene regulation. For example, chromatin contact mapping suggests that like in animals, gene expression can be influenced by multiple regulatory regions and that conversely, an individual regulatory region can modulate multiple genes (Wang et al., 2017; Ricci et al., 2019; Gasperini et al., 2020).

Understanding this complexity will likely shed light on prior genetic data and assist with future engineering efforts.

## PROSPECTS FOR MINING REGULATORY DIVERSITY IN EXISTING GERMPLASM

*De novo* whole genome assembly is becoming wide available opening the door for mining regulatory diversity among not only many different plant species, but also closely related inbred lines, accessions, and varieties (Tao et al., 2019; Danilevicz et al., 2020). Such pan-genome collections allow for identification of regulatory variants including both coding and expression alleles including those associated with gene presence/absence, copy number variation, SNPs, indels, and structural variation, and are likely to be highly informative (Darracq et al., 2018; Sun et al., 2018; Gao et al., 2019; Yang et al., 2019a,b; Zhou et al., 2019; Alonge et al., 2020; Song et al., 2020). Similarly, understanding regulatory divergence among sub-genomes in polyploidy species is another exciting yet challenging prospect (Bao et al., 2019). Annotation of both conserved and accession-specific functional elements within these assemblies will likely require both empirical and machine learning based techniques (Michael and VanBuren, 2020). Among these annotation efforts, cataloging and characterizing CREs and individual TF binding events in plant genomes will be essential for understanding transcriptional and phenotypic variation. Much like the genetic maps and gene maps that have guided plant molecular genetics research for the past several decades, we envision that physical maps of annotated non-coding regulatory regions and CREs will be highly useful for both basic research and precision plant breeding. The generation of species-specific "genomic navigation systems" could transform research in much the same way that cellular navigation systems have enabled expanded and more efficient travel in everyday life. Ultimately, the ability to use CRISPR-based technologies to edit specific regulatory elements and alter transcriptional outputs offers great promise for engineering desirable traits (Rodríguez-Leal et al., 2017; Eshed and Lippman, 2019), providing new ways to increase genetic gain and affording a broader spectrum of genetic variation than what is seen in nature, ultimately transforming our approach to crop improvement.

## AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

## FUNDING

MG, FF, and AG are supported by a grant from the National Science Foundation (TRTech-PGR IOS1916804).

## REFERENCES

Alonge, M., Wang, X., Benoit, M., Soyk, S., Pereira, L., Zhang, L., et al. (2020). Major impacts of widespread structural variation on gene expression and crop improvement in tomato. *Cell* 182, 145.e23–161.e23. doi: 10.1016/j.cell.2020.05.021

Andersson, R., and Sandelin, A. (2020). Determinants of enhancer and promoter activities of regulatory elements. *Nat. Rev. Genet.* 21, 71–87. doi: 10.1038/s41576-019-0173-8

Arnold, C. D., Gerlach, D., Stelzer, C., Boryń, Ł. M., Rath, M., and Stark, A. (2013). Genome-wide quantitative enhancer activity maps identified by STARR-seq. *Science* 339, 1074–1077. doi: 10.1126/science.1232542

Bao, Y., Hu, G., Grover, C. E., Conover, J., Yuan, D., and Wendel, J. F. (2019). Unraveling cis and trans regulatory evolution during cotton domestication. *Nat. Commun.* 10:5399. doi: 10.1038/s41467-019-13386-w

Bartlett, A., O'Malley, R. C., Huang, S. C., Galli, M., Nery, J. R., Gallavotti, A., et al. (2017). Mapping genome-wide transcription-factor binding sites using DAP-seq. *Nat. Protoc.* 12, 1659–1672. doi: 10.1038/nprot.2017.055

Betsiashvili, M., Ahern, K. R., and Jander, G. (2015). Additive effects of two quantitative trait loci that confer *Rhopalosiphum maidis* (corn leaf aphid) resistance in maize inbred line Mo17. *J. Exp. Bot.* 66, 571–578. doi: 10.1093/jxb/eru379

Bolduc, N., Yilmaz, A., Mejia-Guerra, M. K., Morohashi, K., O'Connor, D., Grotewold, E., et al. (2012). Unraveling the KNOTTED1 regulatory network in maize meristems. *Genes Dev.* 26, 1685–1690. doi: 10.1101/gad.193433.112

Brooks, M. D., Cirrone, J., Pasquino, A. V., Alvarez, J. M., Swift, J., Mittal, S., et al. (2019). Network walking charts transcriptional dynamics of nitrogen signaling by integrating validated and predicted genome-wide interactions. *Nat. Commun.* 10:1569. doi: 10.1038/s41467-019-09522-1

Butelli, E., Licciardello, C., Zhang, Y., Liu, J., Mackay, S., Bailey, P., et al. (2012). Retrotransposons control fruit-specific, cold-dependent accumulation of anthocyanins in blood oranges. *Plant Cell* 24, 1242–1255. doi: 10.1105/tpc.111.095232

Castelletti, S., Tuberosa, R., Pindo, M., and Salvi, S. (2014). A MITE transposon insertion is associated with differential methylation at the maize flowering time QTL vgt1. *G3 (Bethesda)* 4, 805–812. doi: 10.1534/g3.114.010686

Chen, K., Wang, Y., Zhang, R., Zhang, H., and Gao, C. (2019). CRISPR/Cas genome editing and precision plant breeding in agriculture. *Annu. Rev. Plant Biol.* 70, 667–697. doi: 10.1146/annurev-arplant-050718-100049

Concia, L., Veluchamy, A., Ramirez-Prado, J. S., Martin-Ramirez, A., Huang, Y., Perez, M., et al. (2020). Wheat chromatin architecture is organized in genome territories and transcription factories. *Genome Biol.* 21:104. doi: 10.1186/s13059-020-01998-1

Crisp, P. A., Marand, A. P., Noshay, J. M., Zhou, P., Lu, Z., Schmitz, R. J., et al. (2020). Stable unmethylated DNA demarcates expressed genes and their cis-regulatory space in plant genomes. *Proc. Natl. Acad. Sci. U.S.A.* 117, 23991–24000. doi: 10.1073/pnas.2010250117

Crisp, P. A., Noshay, J. M., Anderson, S. N., and Springer, N. M. (2019). Opportunities to use DNA methylation to distil functional elements in large crop genomes. *Mol. Plant* 12, 282–284. doi: 10.1016/j.molp.2019.02.006

Danilevicz, M. F., Tay Fernandez, C. G., Marsh, J. I., Bayer, P. E., and Edwards, D. (2020). Plant pangenomics: approaches, applications and advancements. *Curr. Opin. Plant Biol.* 54, 18–25. doi: 10.1016/j.pbi.2019.12.005

Darracq, A., Vitte, C., Nicolas, S., Duarte, J., Pichon, J. P., Mary-Huard, T., et al. (2018). Sequence analysis of European maize inbred line F2 provides new insights into molecular and chromosomal characteristics of presence/absence variants. *BMC Genomics* 19:119. doi: 10.1186/s12864-018-4490-7

de Boer, C. G., Vaishnav, E. D., Sadeh, R., Abeyta, E. L., Friedman, N., and Regev, A. (2020). Deciphering eukaryotic gene-regulatory logic with 100 million random promoters. *Nat. Biotechnol.* 38, 56–65. doi: 10.1038/s41587-019-0315-8

Deplancke, B., Alpern, D., and Gardeux, V. (2016). The genetics of transcription factor DNA binding variation. *Cell* 166, 538–554. doi: 10.1016/j.cell.2016.07.012

Dong, Z., Alexander, M., and Chuck, G. (2019). Understanding grass domestication through maize mutants. *Trends Genet.* 35, 118–128. doi: 10.1016/j.tig.2018.10.007

Dong, P., Tu, X., Chu, P. Y., Lü, P., Zhu, N., Grierson, D., et al. (2017). 3D chromatin architecture of large plant genomes determined by local a/B compartments. *Mol. Plant* 10, 1497–1509. doi: 10.1016/j.molp.2017.11.005

Eichten, S. R., Ellis, N. A., Makarevitch, I., Yeh, C. T., Gent, J. I., Guo, L., et al. (2012). Spreading of heterochromatin is limited to specific families of maize retrotransposons. *PLoS Genet.* 8:e1003127. doi: 10.1371/journal.pgen.1003127

Eshed, Y., and Lippman, Z. B. (2019). Revolutions in agriculture chart a course for targeted breeding of old and new crops. *Science* 366:eaax0025. doi: 10.1126/science.aax0025

Feng, F., Qi, W., Lv, Y., Yan, S., Xu, L., Yang, W., et al. (2018). OPAQUE11 is a central hub of the regulatory network for maize endosperm development and nutrient metabolism. *Plant Cell* 30, 375–396. doi: 10.1105/tpc.17.00616

Fullwood, M. J., Liu, M. H., Pan, Y. F., Liu, J., Xu, H., Mohamed, Y. B., et al. (2009). An oestrogen-receptor-α-bound human chromatin interactome. *Nature* 462, 58–64. doi: 10.1038/nature08497

Galli, M., Khakhar, A., Lu, Z., Chen, Z., Sen, S., Joshi, T., et al. (2018). The DNA binding landscape of the maize AUXIN RESPONSE FACTOR family. *Nat. Commun.* 9:4526. doi: 10.1038/s41467-018-06977-6

Gao, L., Gonda, I., Sun, H., Ma, Q., Bao, K., Tieman, D. M., et al. (2019). The tomato pan-genome uncovers new genes and a rare allele regulating fruit flavor. *Nat. Genet.* 51, 1044–1051. doi: 10.1038/s41588-019-0410-2

Gasperini, M., Tome, J. M., and Shendure, J. (2020). Towards a comprehensive catalogue of validated and target-linked human enhancers. *Nat. Rev. Genet.* 21, 292–310. doi: 10.1038/s41576-019-0209-0

Gregis, V., Andrés, F., Sessa, A., Guerra, R. F., Simonini, S., Mateos, J. L., et al. (2013). Identification of pathways directly regulated by SHORT VEGETATIVE PHASE during vegetative and reproductive development in *Arabidopsis*. *Genome Biol.* 14:R56. doi: 10.1186/gb-2013-14-6-r56

Han, J. J., Jackson, D., and Martienssen, R. (2012). Pod corn is caused by rearrangement at the Tunicate1 locus. *Plant Cell* 24, 2733–2744. doi: 10.1105/tpc.112.100537

Hardison, R. C., and Taylor, J. (2012). Genomic approaches towards finding cis-regulatory modules in animals. *Nat. Rev. Genet.* 13, 469–483. doi: 10.1038/nrg3242

Hernandez, J. M., Feller, A., Morohashi, K., Frame, K., and Grotewold, E. (2007). The basic helix-loop-helix domain of maize R links transcriptional regulation and histone modifications by recruitment of an EMSY-related factor. *Proc. Natl. Acad. Sci. U. S. A.* 104, 17222–17227. doi: 10.1073/pnas.0705629104

Hirsch, C. D., and Springer, N. M. (2017). Transposable element influences on gene expression in plants. *Biochim. Biophys. Acta Gene Regul. Mech.* 1860, 157–165. doi: 10.1016/j.bbagrm.2016.05.010

Huang, S. -S. C., and Ecker, J. R. (2018). Piecing together cis-regulatory networks: insights from epigenomics studies in plants. *Wiley Interdiscip. Rev. Syst. Biol. Med.* 10:e1411. doi: 10.1002/wsbm.1411

Huang, W., Pérez-García, P., Pokhilko, A., Millar, A. J., Antoshechkin, I., Riechmann, J. L., et al. (2012). Mapping the core of the *Arabidopsis* circadian clock defines the network structure of the oscillator. *Science* 336, 75–79. doi: 10.1126/science.1219075

Huang, C., Sun, H., Xu, D., Chen, Q., Liang, Y., Wang, X., et al. (2017). ZmCCT9 enhances maize adaptation to higher latitudes. *Proc. Natl. Acad. Sci. U. S. A.* 115, E334–E341. doi: 10.1073/pnas.1718058115

Jo, L., Pelletier, J. M., Hsu, S. W., Baden, R., Goldberg, R. B., and Harada, J. J. (2020). Combinatorial interactions of the LEC1 transcription factor specify diverse developmental programs during soybean seed development. *Proc. Natl. Acad. Sci. U. S. A.* 117, 1223–1232. doi: 10.1073/pnas.1918441117

Johnson, D. S., Mortazavi, A., Myers, R. M., and Wold, B. (2007). Genome-wide mapping of in vivo-DNA interactions. *Science* 316, 1497–1502. doi: 10.1126/science.1141319

Jolma, A., Yan, J., Whitington, T., Toivonen, J., Nitta, K. R., Rastas, P., et al. (2013). DNA-binding specificities of human transcription factors. *Cell* 152, 327–339. doi: 10.1016/j.cell.2012.12.009

Jores, T., Tonnies, J., Dorrity, M. W., Cuperus, J., Fields, S., and Queitsch, C. (2020). Identification of plant enhancers and their constituent elements by STARR-seq in tobacco leaves. *Plant Cell* 32, 2120–2131. doi: 10.1105/tpc.20.00155

Jung, J. -H., Domijan, M., Klose, C., Biswas, S., Ezer, D., Gao, M., et al. (2016). Phytochromes function as thermosensors in *Arabidopsis*. *Science* 354, 886–889. doi: 10.1126/science.aaf6005

Kaufmann, K., Muiño, J. M., Østerås, M., Farinelli, L., Krajewski, P., and Angenent, G. C. (2010). Chromatin immunoprecipitation (ChIP) of plant transcription factors followed by sequencing (ChIP-SEQ) or hybridization to whole genome arrays (ChIP-CHIP). *Nat. Protoc.* 5, 457–472. doi: 10.1038/nprot.2009.244

Kidder, B. L., Hu, G., and Zhao, K. (2011). ChIP-seq: technical considerations for obtaining high-quality data. *Nat. Immunol.* 12, 918–922. doi: 10.1038/ni.2117

Klemm, S. L., Shipony, Z., and Greenleaf, W. J. (2019). Chromatin accessibility and the regulatory epigenome. *Nat. Rev. Genet.* 20, 207–220. doi: 10.1038/s41576-018-0089-8

Kobayashi, S., Goto-Yamamoto, N., and Hirochika, H. (2004). Retrotransposon-induced mutations in grape skin color. *Science* 304:982. doi: 10.1126/science.1095011

Kong, Q., Pattanaik, S., Feller, A., Werkman, J. R., Chai, C., Wang, Y., et al. (2012). Regulatory switch enforced by basic helix-loop-helix and ACT-domain mediated dimerizations of the maize transcription factor R. *Proc. Natl. Acad. Sci.* 109, E2091–E2097. doi: 10.1073/pnas.1205513109

Konishi, S., Izawa, T., Lin, S. Y., Ebana, K., Fukuta, Y., Sasaki, T., et al. (2006). An SNP caused loss of seed shattering during rice domestication. *Science* 312, 1392–1396. doi: 10.1126/science.1126410

Landt, S. G., Marinov, G. K., Kundaje, A., Kheradpour, P., Pauli, F., Batzoglou, S., et al. (2012). ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. *Genome Res.* 22, 1813–1831. doi: 10.1101/gr.136184.111

Lau, O. S., Davies, K. A., Chang, J., Adrian, J., Rowe, M. H., Ballenger, C. E., et al. (2014). Direct roles of SPEECHLESS in the specification of stomatal self-renewing cells. *Science* 345, 1605–1609. doi: 10.1126/science.1256888

Lee, J. H., Jin, S., Kim, S. Y., Kim, W., and Ahn, J. H. (2017). A fast, efficient chromatin immunoprecipitation method for studying protein-DNA binding in *Arabidopsis* mesophyll protoplasts. *Plant Methods* 13:42. doi: 10.1186/s13007-017-0192-4

Li, E., Liu, H., Huang, L., Zhang, X., Dong, X., Song, W., et al. (2019). Long-range interactions between proximal and distal regulatory regions in maize. *Nat. Commun.* 10:2633. doi: 10.1038/s41467-019-10603-4

Li, C., Qiao, Z., Qi, W., Wang, Q., Yuan, Y., Yang, X., et al. (2015). Genome-wide characterization of *cis*-acting DNA targets reveals the transcriptional regulatory framework of *Opaque2* in maize. *Plant Cell* 27, 532–545. doi: 10.1105/tpc.114.134858

Liu, C., Cheng, Y. J., Wang, J. W., and Weigel, D. (2017). Prominent topologically associated domains differentiate global chromatin packing in rice from *Arabidopsis*. *Nat. Plants* 3, 742–748. doi: 10.1038/s41477-017-0005-9

Liu, W., and Stewart, C. N. (2016). Plant synthetic promoters and transcription factors. *Curr. Opin. Biotechnol.* 37, 36–44. doi: 10.1016/j.copbio.2015.10.001

Lu, Z., Marand, A. P., Ricci, W. A., Ethridge, C. L., Zhang, X., and Schmitz, R. J. (2019). The prevalence, evolution and chromatin signatures of plant regulatory elements. *Nat. Plants* 5, 1250–1259. doi: 10.1038/s41477-019-0548-z

Lü, P., Yu, S., Zhu, N., Chen, Y. R., Zhou, B., Pan, Y., et al. (2018). Genome encode analyses reveal the basis of convergent evolution of fleshy fruit ripening. *Nat. Plants* 4, 784–791. doi: 10.1038/s41477-018-0249-z

Maher, K. A., Bajic, M., Kajala, K., Reynoso, M., Pauluzzi, G., West, D. A., 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. doi: 10.1105/tpc.17.00581

Manning, K., Tör, M., Poole, M., Hong, Y., Thompson, A. J., King, G. J., et al. (2006). A naturally occurring epigenetic mutation in a gene encoding an SBP-box transcription factor inhibits tomato fruit ripening. *Nat. Genet.* 38, 948–952. doi: 10.1038/ng1841

Marand, A. P., Zhang, T., Zhu, B., and Jiang, J. (2017). Towards genome-wide prediction and characterization of enhancers in plants. *Biochim. Biophys. Acta Gene Regul. Mech.* 1860, 131–139. doi: 10.1016/j.bbagrm.2016.06.006

Martin, A., Troadec, C., Boualem, A., Rajab, M., Fernandez, R., Morin, H., et al. (2009). A transposon-induced epigenetic change leads to sex determination in melon. *Nature* 461, 1135–1138. doi: 10.1038/nature08498

Mascher, M., Gundlach, H., Himmelbach, A., Beier, S., Twardziok, S. O., Wicker, T., et al. (2017). A chromosome conformation capture ordered sequence of the barley genome. *Nature* 544, 427–433. doi: 10.1038/nature22043

Meyer, R. S., and Purugganan, M. D. (2013). Evolution of crop species: genetics of domestication and diversification. *Nat. Rev. Genet.* 14, 840–852. doi: 10.1038/nrg3605

Michael, T. P., and VanBuren, R. (2020). Building near-complete plant genomes. *Curr. Opin. Plant Biol.* 54, 26–33. doi: 10.1016/j.pbi.2019.12.009

Mumbach, M., Rubin, A., Flynn, R., Dai, C., Khavari, P., Greenleaf, W., et al. (2016). HiChIP: efficient and sensitive analysis of protein-directed genome architecture. *Nat. Methods* 13, 919–922. doi: 10.1038/nmeth.3999

Noshay, J. M., Marand, A. P., Anderson, S. N., Zhou, P., Guerra, M. K. M., Lu, Z., et al. (2020). Cis-regulatory elements within TEs can influence expression of nearby maize genes. bioRxiv [Preprint]. doi: 10.1101/2020.05.20.107169

O'Malley, R. C., Huang, S. S. C., Song, L., Lewsey, M. G., Bartlett, A., Nery, J. R., et al. (2016). Cistrome and epicistrome features shape the regulatory DNA landscape. *Cell* 165, 1280–1292. doi: 10.1016/j.cell.2016.04.038

Oka, R., Zicola, J., Weber, B., Anderson, S. N., Hodgman, C., Gent, J. I., et al. (2017). Genome-wide mapping of transcriptional enhancer candidates using DNA and chromatin features in maize. *Genome Biol.* 18:137. doi: 10.1186/s13059-017-1273-4

Olsen, K. M., and Wendel, J. F. (2013). Crop plants as models for understanding plant adaptation and diversification. *Front. Plant Sci.* 4:290. doi: 10.3389/fpls.2013.00290

Para, A., Li, Y., Marshall-Colón, A., Varala, K., Francoeur, N. J., Moran, T. M., et al. (2014). Hit-and-run transcriptional control by bZIP1 mediates rapid nutrient signaling in *Arabidopsis*. *Proc. Natl. Acad. Sci. U. S. A.* 111, 10371–10376. doi: 10.1073/pnas.1404657111

Park, P. J. (2009). ChIP-seq: advantages and challenges of a maturing technology. *Nat. Rev. Genet.* 10, 669–680. doi: 10.1038/nrg2641

Parvathaneni, R. K., Bertolini, E., Shamimuzzaman, M., Vera, D. L., Lung, P. -Y., Rice, B. R., et al. (2020). The regulatory landscape of early maize inflorescence development. *Genome Biol.* 21:165. doi: 10.1186/s13059-020-02070-8

Pautler, M., Eveland, A. L., Larue, T., Yang, F., Weeks, R., Lunde, C., et al. (2015). FASCIATED EAR4 encodes a bZIP transcription factor that regulates shoot meristem size in maize. *Plant Cell* 27, 104–120. doi: 10.1105/tpc.114.132506

Peng, Y., Xiong, D., Zhao, L., Ouyang, W., Wang, S., Sun, J., et al. (2019). Chromatin interaction maps reveal genetic regulation for quantitative traits in maize. *Nat. Commun.* 10:2632. doi: 10.1038/s41467-019-10602-5

Pliner, H. A., Packer, J. S., Steemers, F. J., Shendure, J., Aghamirzaie, D., Srivatsan, S., et al. (2018). Cicero predicts cis-regulatory DNA interactions from single-cell chromatin accessibility data technology. *Mol. Cell* 71, 858.e8–871.e8. doi: 10.1016/j.molcel.2018.06.044

Ricci, W. A., Lu, Z., Ji, L., Marand, A. P., Ethridge, C. L., Murphy, N. G., et al. (2019). Widespread long-range cis-regulatory elements in the maize genome. *Nat. Plants* 5, 1237–1249. doi: 10.1038/s41477-019-0547-0

Rodgers-Melnick, E., Vera, D. L., Bass, H. W., and Buckler, E. S. (2016). Open chromatin reveals the functional maize genome. *Proc. Natl. Acad. Sci.* 113, E3177–E3184. doi: 10.1073/pnas.1525244113

Rodríguez-Leal, D., Lemmon, Z. H., Man, J., Bartlett, M. E., and Lippman, Z. B. (2017). Engineering quantitative trait variation for crop improvement by genome editing. *Cell* 171, 470.e8–480.e8. doi: 10.1016/j.cell.2017.08.030

Rossi, M., Duffy, T., Conti, G., Almeida, J., Bermudez, L., Fernie, A. R., et al. (2014). Natural occurring epialleles determine vitamin E accumulation in tomato fruits. *Nat. Commun.* 5:3027. doi: 10.1038/ncomms5027

Rowley, M. J., Nichols, M. H., Lyu, X., Ando-Kuri, M., Rivera, I. S. M., Hermetz, K., et al. (2017). Evolutionarily conserved principles predict 3D chromatin organization. *Mol. Cell* 67, 837.e7–852.e7. doi: 10.1016/j.molcel.2017.07.022

Shlyueva, D., Stampfel, G., and Stark, A. (2014). Transcriptional enhancers: from properties to genome-wide predictions. *Nat. Rev. Genet.* 15, 272–286. doi: 10.1038/nrg3682

Skene, P. J., and Henikoff, S. (2017). An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites. *eLife* 6:e21856. doi: 10.7554/eLife.21856

Skene, P. J., Henikoff, J. G., and Henikoff, S. (2018). Targeted in situ genome-wide profiling with high efficiency for low cell numbers. *Nat. Protoc.* 13, 1006–1019. doi: 10.1038/nprot.2018.015

Slattery, M., Zhou, T., Yang, L., Dantas Machado, A. C., Gordân, R., and Rohs, R. (2014). Absence of a simple code: how transcription factors read the genome. *Trends Biochem. Sci.* 39, 381–399. doi: 10.1016/j.tibs.2014.07.002

Song, J. M., Guan, Z., Hu, J., Guo, C., Yang, Z., Wang, S., et al. (2020). Eight high-quality genomes reveal pan-genome architecture and ecotype differentiation of *Brassica napus*. *Nat. Plants* 6, 34–45. doi: 10.1038/s41477-019-0577-7

Song, L., Huang, S. -S. C., Wise, A., Castanon, R., Nery, J. R., Chen, H., et al. (2016). A transcription factor hierarchy defines an environmental stress response network. *Science* 354:aag1550. doi: 10.1126/science.aag1550

Spitz, F., and Furlong, E. E. M. (2012). Transcription factors: from enhancer binding to developmental control. *Nat. Rev. Genet.* 13, 613–626. doi: 10.1038/nrg3207

Springer, N., de León, N., and Grotewold, E. (2019). Challenges of translating gene regulatory information into agronomic improvements. *Trends Plant Sci.* 24, 1075–1082. doi: 10.1016/j.tplants.2019.07.004

Stam, M., Belele, C., Dorweiler, J. E., and Chandler, V. L. (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. doi: 10.1101/gad.1006702

Studer, A., Zhao, Q., Ross-Ibarra, J., and Doebley, J. (2011). Identification of a functional transposon insertion in the maize domestication gene tb1. *Nat. Genet.* 43, 1160–1163. doi: 10.1038/ng.942

Sullivan, A. M., Arsovski, A. A., Lempe, J., Bubb, K. L., Weirauch, M. T., Sabo, P. J., et al. (2014). Mapping and dynamics of regulatory DNA and transcription factor networks in *A. thaliana*. *Cell Rep.* 8, 2015–2030. doi: 10.1016/j.celrep.2014.08.019

Sun, Y., Dong, L., Zhang, Y., Lin, D., Xu, W., Ke, C., et al. (2020). 3D genome architecture coordinates trans and cis regulation of differentially expressed ear and tassel genes in maize. *Genome Biol.* 21:143. doi: 10.1186/s13059-020-02063-7

Sun, S., Zhou, Y., Chen, J., Shi, J., Zhao, H., Zhao, H., et al. (2018). Extensive intraspecific gene order and gene structural variations between Mo17 and other maize genomes. *Nat. Genet.* 50, 1289–1295. doi: 10.1038/s41588-018-0182-0

Sundaram, V., Cheng, Y., Ma, Z., Li, D., Xing, X., Edge, P., et al. (2014). Widespread contribution of transposable elements to the innovation of gene regulatory networks. *Genome Res.* 24, 1963–1976. doi: 10.1101/gr.168872.113

Swinnen, G., Goossens, A., and Pauwels, L. (2016). Lessons from domestication: targeting cis-regulatory elements for crop improvement. *Trends Plant Sci.* 21, 506–515. doi: 10.1016/j.tplants.2016.01.014

Tao, Y., Zhao, X., Mace, E., Henry, R., and Jordan, D. (2019). Exploring and exploiting pan-genomics for crop improvement. *Mol. Plant* 12, 156–169. doi: 10.1016/j.molp.2018.12.016

Todeschini, A. L., Georges, A., and Veitia, R. A. (2014). Transcription factors: specific DNA binding and specific gene regulation. *Trends Genet.* 30, 211–219. doi: 10.1016/j.tig.2014.04.002

Tsuda, K., Kurata, N., Ohyanagi, H., and Hake, S. (2014). Genome-wide study of *KNOX* regulatory network reveals brassinosteroid catabolic genes important for shoot meristem function in rice. *Plant Cell* 26, 3488–3500. doi: 10.1105/tpc.114.129122

Tu, X., Mejía-Guerra, M. K., Franco, J. A. V., Tzeng, D., Chu, P.-Y., Dai, X., et al. (2020). The transcription regulatory code of a plant leaf. *Nat. Commun.* 11:5089. doi: 10.1038/s41467-020-18832-8

van Steensel, B., and Dekker, J. (2010). Genomics tools for unraveling chromosome architecture. *Nat. Biotechnol.* 28, 1089–1095. doi: 10.1038/nbt.1680

Wang, S., Li, S., Liu, Q., Wu, K., Zhang, J., Wang, S., et al. (2015b). The OsSPL16-GW7 regulatory module determines grain shape and simultaneously improves rice yield and grain quality. *Nat. Genet.* 47, 949–954. doi: 10.1038/ng.3352

Wang, C., Liu, C., Roqueiro, D., Grimm, D., Schwab, R., Becker, C., et al. (2015a). Genome-wide analysis of local chromatin packing in *Arabidopsis thaliana*. *Genome Res.* 25, 246–256. doi: 10.1101/gr.170332.113

Wang, M., Tu, L., Lin, M., Lin, Z., Wang, P., Yang, Q., et al. (2017). Asymmetric subgenome selection and cis-regulatory divergence during cotton domestication. *Nat. Genet.* 49, 579–587. doi: 10.1038/ng.3807

Wang, X., Wang, H., Liu, S., Ferjani, A., Li, J., Yan, J., et al. (2016). Genetic variation in ZmVPP1 contributes to drought tolerance in maize seedlings. *Nat. Genet.* 48, 1233–1241. doi: 10.1038/ng.3636

Weirauch, M. T., Yang, A., Albu, M., Cote, A. G., Montenegro-Montero, A., Drewe, P., et al. (2014). Determination and inference of eukaryotic transcription factor sequence specificity. *Cell* 158, 1431–1443. doi: 10.1016/j.cell.2014.08.009

Yang, Z., Ge, X., Yang, Z., Qin, W., Sun, G., Wang, Z., et al. (2019b). Extensive intraspecific gene order and gene structural variations in upland cotton cultivars. *Nat. Commun.* 10:2989. doi: 10.1038/s41467-019-10820-x

Yang, Q., Li, Z., Li, W., Ku, L., Wang, C., Ye, J., et al. (2013). CACTA-like transposable element in ZmCCT attenuated photoperiod sensitivity and accelerated the postdomestication spread of maize. *Proc. Natl. Acad. Sci. U. S. A.* 110, 16969–16974. doi: 10.1073/pnas.1310949110

Yang, N., Liu, J., Gao, Q., Gui, S., Chen, L., Yang, L., et al. (2019a). Genome assembly of a tropical maize inbred line provides insights into structural variation and crop improvement. *Nat. Genet.* 51, 1052–1059. doi: 10.1038/s41588-019-0427-6

Yoo, S. D., Cho, Y. H., and Sheen, J. (2007). *Arabidopsis* mesophyll protoplasts: a versatile cell system for transient gene expression analysis. *Nat. Protoc.* 2, 1565–1572. doi: 10.1038/nprot.2007.199

Zhao, H., Zhang, W., Chen, L., Wang, L., Marand, A. P., Wu, Y., et al. (2018). Proliferation of regulatory DNA elements derived from transposable elements in the maize genome. *Plant Physiol.* 176, 2789–2803. doi: 10.1104/pp.17.01467

Zheng, L., McMullen, M. D., Bauer, E., Schön, C. C., Gierl, A., and 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. doi: 10.1093/jxb/erv192

Zhong, S., Fei, Z., Chen, Y. R., Zheng, Y., Huang, M., Vrebalov, J., et al. (2013). Single-base resolution methylomes of tomato fruit development reveal epigenome modifications associated with ripening. *Nat. Biotechnol.* 31, 154–159. doi: 10.1038/nbt.2462

Zhou, Y., Minio, A., Massonnet, M., Solares, E., Lv, Y., Beridze, T., et al. (2019). The population genetics of structural variants in grapevine domestication. *Nat. Plants* 5, 965–979. doi: 10.1038/s41477-019-0507-8

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

*Copyright © 2020 Galli, Feng and Gallavotti. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.*