



Contents lists available at ScienceDirect

Biochimica et Biophysica Acta

journal homepage: www.elsevier.com/locate/bbagrm

Towards genome-wide prediction and characterization of enhancers in plants[☆]

Alexandre P. Marand, Tao Zhang, Bo Zhu, Jiming Jiang^{*}

Department of Horticulture, University of Wisconsin-Madison, Madison, WI 53706, USA

ARTICLE INFO

Article history:

Received 12 April 2016

Received in revised form 13 June 2016

Accepted 14 June 2016

Available online xxx

Keywords:

Enhancer

Promoter

Transcription co-factor

Open chromatin

Histone modification

ABSTRACT

Enhancers are important *cis*-regulatory DNA elements that regulate transcription programs by recruiting transcription factors and directing them to the promoters of target genes in a cell-type/tissue-specific manner. The expression of a gene can be regulated by one or multiple enhancers at different developmental stages and/or in different tissues. Enhancers are difficult to identify because of their unpredictable positions relative to their cognate promoters. Remarkably, only a handful of enhancers have been identified in plant species largely due to the lack of general approaches for enhancer identification. Extensive genomic and epigenomic research in mammalian species has revealed that the genomic locations of enhancers can be predicted based on the binding sites of transcriptional co-factors and several distinct features associated with open chromatin. Here we review the methodologies used in enhancer prediction in mammalian species. We also review the recent applications of these methodologies in *Arabidopsis thaliana* and discuss the future directions of enhancer identification in plants. This article is part of a Special Issue entitled: Plant Gene Regulatory Mechanisms and Networks, edited by Dr. Erich Grotewold and Dr. Nathan Springer.

© 2016 Elsevier B.V. All rights reserved.

1. Introduction

Plant development and responses to environmental and hormonal cues are carefully regulated by precise orchestration of gene transcription programs [1]. The spatiotemporal specificity of gene transcription can be largely attributed to the binding of transcription factors (TFs) to regulatory DNA elements [2]. Generally, the presence of regulatory elements located proximal to the transcription start site (TSS), also known as the core promoter, is sufficient for the assembly of transcriptional machinery, often leading to low or moderate gene expression. However, the participation of some promoter-distal regulatory elements can drastically alter the rate and quantity of mRNA biosynthesis through poorly understood mechanisms. These distant sequences contain clusters of short, 4–30 base pair (bp) DNA motifs which function as binding sites for sequence-specific TFs and collectively are referred to *cis*-regulatory elements (CREs) (Fig. 1). CREs can function in gene transcription, repression, chromatin organization and a multitude of other functions, however, this review will pay particular attention to the transcription activating CREs, termed enhancers.

Enhancer-bound TFs are capable of recruiting co-activators and other TFs that interact with different components of the mediator complex in order to recruit RNA polymerase II (RNAPII) to the TSS [3]. However, for distally activated elements to facilitate the interaction between TFs and RNAPII, the enhancer-associated chromatin must loop over the intervening sequences, and direct interaction between the enhancer-bound TF complex, the target gene promoter, and the transcriptional machinery, in a process that is typically mediated and stabilized by the chromosome-associated multi-subunit protein complex, cohesin [4] (Fig. 1). This demonstrates the ability of enhancers to act independently of distance, position and orientation with respect to the target gene(s). Surprisingly, in a few rare cases, enhancers have been implicated in interchromosomal regulation, by activating transcriptional programs of target genes found on different chromosomes, although the legitimacy of these claims remains somewhat contentious [5,6]. Additionally, enhancer-bound transcription factors are responsible for the recruitment of histone modifying enzymes and ATP-dependent chromatin remodeling complexes, cofactors which aid in the unraveling of tightly condensed chromatin fibers and thus increase the accessibility of DNA to other transcription regulatory proteins [7].

Substantial advancements in computational and molecular biology techniques have enabled the implementation of genome-wide tools for analysis of the genomic and functional characteristics of enhancers. These methods revealed several chromatin features associated with enhancers, such as hypersensitivity to DNase I, the presence of specific

[☆] This article is part of a Special Issue entitled: Plant Gene Regulatory Mechanisms and Networks, edited by Dr. Erich Grotewold and Dr. Nathan Springer.

^{*} Corresponding author.

E-mail address: jjiang1@wisc.edu (J. Jiang).

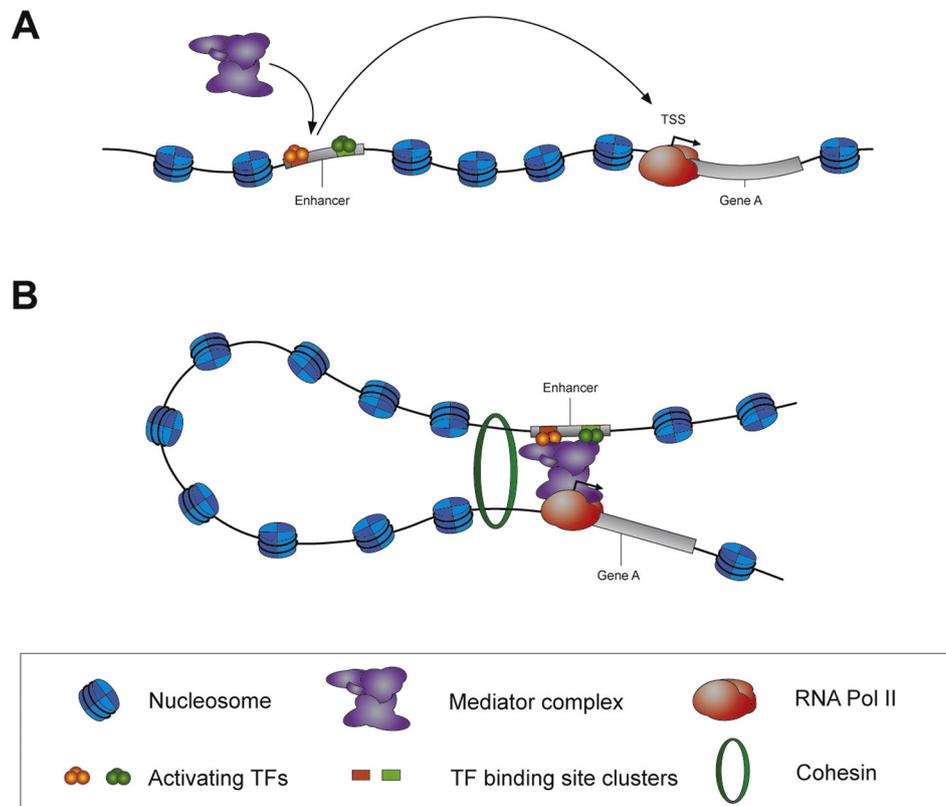


Fig. 1. A model for enhancer regulation of gene expression. Enhancers are distinct genomic regions that contain clusters of transcription factor binding sites which can enhance or up-regulate the transcription rate of a target gene(s), regardless of their location, orientation and distance with respect to the target(s). (A) Binding of transcription factors (TFs) to DNA motifs within the enhancer triggers the recruitment of the mediator complex and stimulates chromatin looping. (B) Cohesin stabilizes the enhancer-promoter interaction.

transcriptional cofactors, various histone modifications, chromatin looping, and nucleosome depletion and phasing. When analyzed congruently, enhancer-associated chromatin features lend an attractive and powerful avenue for the identification of novel enhancers. However, most of our current understanding of enhancer function stem from studies initiated by the Encyclopedia of DNA elements (ENCODE) project [8], which aimed to characterize and identify all functional elements, including enhancers, within the human genome. To date, no such effort has been initiated in any plant species, limiting our understanding of these regulatory elements in the context of the plant cells. Nevertheless, with the wide adoption of high-throughput sequencing methods coupled with enhancer-associated chromatin feature enrichment techniques, it has become more feasible than ever to answer basic questions in plant enhancer biology. This review presents the different methods pioneered by mammalian researchers for the identification of enhancers, as well as the genomic characteristics of enhancers that form the basis of these techniques. We also highlight studies, seminal and current, confirming enhancer activity through functional and genomic signature characterization, and discuss the future directions of enhancer research in plants.

2. Genome-wide features of enhancers

Recent advances in bioinformatics and molecular techniques have facilitated the adoption of genome-wide tools for enhancer prediction and identification. Several key chromatin features and enhancer-associated proteins have been thoroughly dissected in mammalian systems, enabling the translation of these approaches for use in the context of plants. A number of fundamental hallmarks have been used to locate enhancers, which include (i) TF-bound genomic regions; (ii) histone modifications of nucleosomes adjacent to enhancers; (iii) the presence of structural, transcriptional, and cofactor proteins, such as cohesin,

RNAPII and the mediator complex distal to promoters; (iv) transcribed enhancer RNAs (eRNAs), and (v) open chromatin conformations. Here we present and discuss the roles of these chromatin features in enhancer biology as determined in mammalian studies as well as the methodologies used for their discovery. In addition, we evaluate the advantages and limitations of these methodologies in plant enhancer identification.

2.1. Transcriptional co-factors

One strategy to reveal genome-wide enhancers is by identifying their underlying chromatin features and associated proteins. To understand how these features are used to identify enhancers, we need to be familiar with the timing and specific roles of enhancer-associated proteins during enhancer-mediated transcription. Activation of an enhancer is strictly initiated by the binding of transcription factors to their cognate DNA motifs intrinsic to the enhancer sequence [9]. Upon DNA binding, TFs have been demonstrated to be responsible for recruiting the mediator complex, a major component of enhancer-directed transcription [10]. TFs additionally signal for the recruitment of chromatin remodeling complexes (Fig. 1), and histone modifying cofactors such as p300/CBP, proteins which directly regulate chromatin structure and accessibility [11]. Lastly, enhancers associate with transcriptional machinery such as general transcription factors and RNAPII. By mapping the locations of these transcriptional cofactors, several studies have demonstrated the effectiveness of this strategy for enhancer discovery.

A high-throughput method is required to reveal protein-DNA associations accurately and on a genome-wide scale. Chromatin immunoprecipitation (ChIP) with antibodies specific to enhancer-associated cofactors or transcription factors coupled with massively parallel sequencing (ChIP-seq) (Fig. 2A) has become the most common platform for the identification of the protein binding sites genome-wide [12,13]. It is

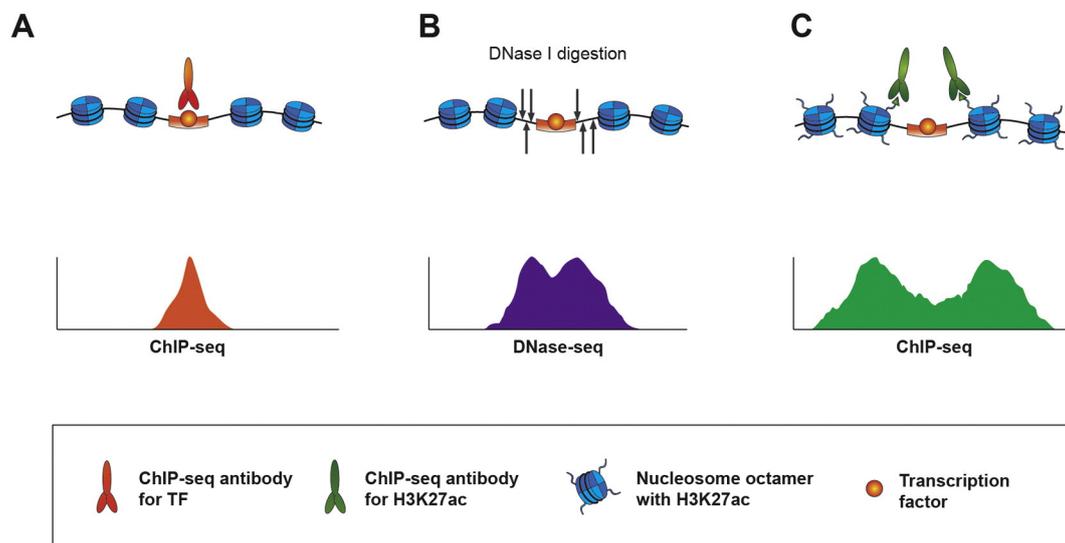


Fig. 2. Common techniques used in identifying chromatin features associated with enhancers. (A) ChIP-seq using antibodies against TFs or transcriptional co-factors. A putative enhancer, representing a protein binding site, is typically identified as single peak centered by a DNA motif recognized by the TF or co-factor. (B) DNase-seq. A putative enhancer, representing a DHS, is typically identified by a “dual peak”, which is centered by a footprint caused by the binding of regulatory protein(s). (C) ChIP-seq using antibodies against specific histone modifications. A putative enhancer, representing a region marked by specific histone modifications, such as H3K27ac, is typically identified by a “dual peak”, which is centered by the enhancer.

known that different tissues exhibit specific transcriptional programs in order to preserve their identity. Visel et al. utilized ChIP-seq of the p300/CPB protein between mouse embryonic, forebrain, midbrain and limb tissues to identify tissue-specific enhancers [14]. Using this strategy, Visel et al. were able to reveal ~5000 distinct tissue-specific enhancers with a high degree of accuracy. This study additionally tested the activity of 122 enhancers by functional characterization of enhancer sequences through the use of enhancer-driven reporter vectors, and further confirmed that ~88% (107/122) of the tested genomic regions demonstrated reproducible reporter activation [14]. Another study in mouse embryonic stem cells examined the genome-wide occupancy of the mediator complex and cohesin as a means of enhancer discovery [15]. The majority of shared mediator and cohesin signals co-localized on active enhancers, a finding which was supplemented by knockdown datasets showing significant transcriptional changes to genes occupied by mediator and cohesin. Collectively, these studies suggest that the presence of enhancer-associated cofactors at a specific locus offers convincing evidence for the positive identification of active enhancers.

2.2. Enhancers and other CREs are found in open chromatin

DNA in all eukaryotic organisms is packaged into chromatin, which is comprised of ~147 bp of DNA tightly wrapped around core histone octamers, known as nucleosomes. Because nucleosomes are composed of positively charged amino acids, they have a high affinity for the negatively charged phosphate backbone of DNA, and due to this strong affinity, genomic regions with a high density of nucleosome occupancy are relatively inaccessible to other DNA binding proteins [16]. CRE sequences must be physically accessible in order to allow TF binding, and because of this requirement, CREs are highly associated with nucleosome-depleted regions (NDRs) or regions of ‘open chromatin’ [17,18]. However, certain pioneer transcription factors can bind to regions of high nucleosome density, or “closed chromatin”, and in turn allow the binding of other TFs, chromatin modifiers and nucleosome remodelers, and thus create an “open chromatin” state at the targeted locus [19]. Though in general, dynamic nucleosome packaging is a direct consequence of regulatory protein binding to CREs in open chromatin, histone acetylation, or activation of nucleosome remodeling complexes, all of which ultimately lead to nucleosome eviction and increased chromatin accessibility at regulatory sites [20,21].

NDRs show a pronounced susceptibility to cleavage by deoxyribonuclease I (DNase I), and are known as DNase I hypersensitive sites (DHSs) [22–26]. Thus, the extent of chromatin accessibility can be directly and reliably measured utilizing DNase I digestion in combination with next generation sequencing (DNase-seq) (Fig. 2B) [22–27]. A plethora of studies in human have utilized this strategy in an attempt to characterize DHS variability across cell-types in order to construct CRE catalogs for specific cells lines. One particular study initiated by the ENCODE project was able to demonstrate that distinct human cell lines exhibit specific DHS profiles, and thus display differential expression programs, leading to identification of genes characteristic to these cell-types [28]. New DHSs arise and disappear between distinct cell-types, suggesting that DHSs play a role in determination and maintenance of cell identity [28]. A large proportion of these DHSs functionally activate gene expression programs, which implies that the overlying CRE in a majority of DHSs function as transcriptional enhancers. Using this information, many labs have exploited DHSs for direct identification of putative enhancer loci. A study of DHSs in mouse brain and neural retina samples revealed that DNase-seq was just as capable at discriminating enhancer loci as p300 and histone 3 lysine 27 acetylation (H3K27ac) ChIP-seq datasets by evaluating regions near candidate genes as well as intergenic regions [29]. Moreover, this particular study successfully demonstrated functional activity of three novel retina-specific enhancers surrounding *Otx2*, a well characterized retina developmental gene, identified through DNase-seq. Thus, systematic evaluation of DNase-seq datasets at intergenic loci coupled with a priori knowledge or gene expression data can allow for identification of CREs, and in particular, putative enhancers and their locations genome-wide.

2.3. Histone modifications associated with enhancers

Following sequence-specific binding of TFs to their cognate motifs within an enhancer, TFs participate in the recruitment of transcriptional cofactors. While cofactors lack DNA-binding capabilities, they offer a range of functional contributions to enhancer activity, such as post-translational histone modifications [30]. Active enhancers are located in regions depleted of nucleosomes, but neighboring nucleosomes often contain histones that are modified post-translationally by transcriptional cofactors with methyl- and acetyltransferase activity (Fig. 2C). Researchers have become interested in defining the chromatin landscape for distinct genomic regions, and with those efforts, several histone-

associated features relevant to enhancer characterization have come to light. Intriguingly, genome-wide studies mapping various histone modifications via ChIP-seq have identified distinctly reproducible patterns of modifications that appear at defined genomic regions, such as active enhancers, promoters and transcriptionally repressed loci [31]. Discerning these discrete histone modification patterns genome-wide can facilitate annotation of genomic enhancers.

Several histone marks have been revealed to be associated with distinct enhancer states, including: (i) histone 3 lysine 4 mono-methylation (H3K4me1); (ii) H3K27ac; (iii) histone 3 lysine 27 tri-methylation (H3K27me3); and (iv) histone 3 lysine 79 tri-methylation (H3K79me3) [32]. Enrichment of some of these modifications at specific genomic sites, H3K4me1 and H3K27ac in particular, are usually strong indicators of transcriptionally active enhancers [33] (Fig. 2C). One particular study in human Treg cells identified enhancers based on the presence of H3K4me1 while utilizing H3K4me3 to mark proximal promoters [34]. The authors additionally cloned a subset of putative enhancers into luciferase reporter constructs and demonstrated that approximately 50% of their candidates had significant transcriptional activation activity. Interestingly, some enhancers may contain H3K4me1, yet may not be active. Studies evaluating the presence of H3K4me1 with either H3K27ac or H3K27me3 have suggested that there may be additional enhancer states which depend on histone modification status in neighboring nucleosomes [35]. H3K27me3 was proposed to play a role in regulating enhancer accessibility to TFs by masking enhancer chromatin, and therefore inhibiting premature TF binding [36]. The reversibility of H3K27me3 is suggestive of dynamic regulation at genomic enhancers, and lends to the hypothesis that H3K27me3 along with H3K4me1 may mark bivalent enhancers, or enhancers which are poised for activation, yet are presently in-active [37]. Thus, the deposition of H3K27ac in place of H3K27me3 can discriminate between active and inactive enhancers carrying H3K4me1 modifications. One such study clearly establishes this hypothesis by evaluating H3K4me1, H3K27ac and H3K27me3 datasets to identify putative poised or active enhancers in five different human cell lines supported by eRNA and gene transcription data sets [35]. The picture of how these modifications aid in enhancer activation and function is only beginning to emerge. However, regardless of the mechanism by which histones become modified, these marks provide a reliable and consistent pattern suggestive of genomic enhancers, particularly when used in combination. Furthermore, by evaluating a host of histone modification marks, systematic classification of active, repressed or poised putative enhancers becomes possible.

2.4. Promoter-enhancer interaction

For an enhancer to actively regulate transcription of its target gene(s), the enhancer must be in close spatial proximity to the promoter(s) of its target(s). The most accepted model for this phenomenon is known as DNA looping (Fig. 1). Strikingly, it remains unclear which factors are responsible for initiating DNA looping, though it has been proposed that these looping structures are in place well in advance of transcription initiation signals [38]. Enhancer loops are known to be stabilized by a multi-subunit chromosome structural protein, cohesin, which is also recognized to play a role in gene repression [15]. It has been suggested that cohesin helps to stabilize enhancer-promoter interactions during enhancer-regulated transcription by scaffolding enhancer and promoter chromatin [39]. The enhancer-associated chromatin harboring TFs, cofactors and transcriptional machinery loop over the intervening sequences in order to permit the interaction between the promoter and enhancer elements (Fig. 1). Recently, several studies have revealed that enhancer-promoter chromatin interaction networks are generally distinct between cell-types, which may be a major reason for cell type-specific transcriptional programs [40]. Furthermore, another study was able to show that in human fibroblast cells, networks of

chromatin interactions are even in place prior to cellular stimulation by an external signal, and interestingly, stimulation does not significantly alter the chromatin architecture [41]. Thus, methods which can identify these interactions between promoters and putative enhancers provide a different and perhaps more direct approach for enhancer identification and validation.

Interactions between distant chromatin segments can be directly measured using chromosome conformation capture (3C) and its variants, including circular chromosome conformation capture (4C), chromosome conformation capture carbon copy (5C), and Hi-C [42]. These methods are based on cross-linking the points of contact between *in vivo* chromatin fragments that are in close spatial proximity within the nucleus. Many studies aimed to discern the mode of interaction between the enhancer chromatin and that of the promoter elements by employing these techniques, and for the most part, have resulted in a general consensus in agreement with the pre-existing DNA-looping model. Although 5C and Hi-C were designed to assess the organizational patterns of the genome, they were not, however, designed specifically to probe the interaction between an enhancer and its target promoter(s) [42]. A recently developed technique based on 3C, chromatin interaction analysis by paired-end tag sequencing (ChIA-PET), essentially couples 3C technology with chromatin immunoprecipitation. Interactions between spatially proximal chromatin segments are preserved and enriched by ChIP; this allows the investigation of chromatin fragments associated with a specific TF or cofactor, which are not only physically bound to DNA, but also in close contact within the nucleus via DNA looping. A recent study utilizing this technique focused on enriching interactions associated with RNAPII [40]. This strategy provides a logical methodology to identify not only active enhancers, but their target promoter(s) as well. However, ChIA-PET is a particularly complex and technically challenging technique and has yet to be realized in plant species owing to these obstacles. Further optimization of ChIA-PET will be necessary before it can become feasible in plants.

2.5. Transcription of genomic enhancers

Recent genomic advances have enabled the identification of novel non-coding transcripts derived from genomic enhancers (eRNAs). Transcription of enhancers was first discovered in 1992 but dissection of their function was not instigated until the late 2000s [43]. Kim et al. were able to demonstrate the presence of eRNAs at enhancer loci by sequencing total RNA (RNA-seq) and positively mapping these non-coding transcripts to enhancer regions bound by RNAPII via ChIP-seq [44]. eRNA quantity positively correlated with mRNA biosynthesis of nearby genes, resulting in speculation that these elements contribute functionally to enhancer activity [44]. Additionally, when the target promoter of an enhancer was knocked out, this subsequently abolished eRNA transcription, prompting the hypothesis that eRNAs may facilitate enhancer-promoter interactions [45]. Though this claim remains somewhat of a controversy as another study indicated that eRNAs may not be relevant to DNA looping since knocking down eRNAs from active enhancers did not affect promoter-enhancer interactions [46]. A more recent examination of eRNAs have advocated for a functional eRNA model in which there are two enhancer states, based on the presence or absence of eRNAs [47]. Evidence for functional eRNAs stem from studies which revealed that by knocking down eRNAs, expression levels of nearby genes were subsequently reduced [46,48,49]. Even more striking, was the discovery that for some enhancers, eRNA sequence is critical to their function, since modifications to eRNA sequences negatively alters the expression of their target genes [50]. Additional evidence for a role of eRNAs in enhancer function was demonstrated by Cheng et al. who determined the existence of two possible states of enhancers, transcribing and non-transcribing, where transcribed enhancers correlated with higher expressed target genes in contrast to non-transcribed enhancers [47]. In addition, transcribed enhancers were more associated with genes implicated in cell type-specific

processes [47]. It remains unclear how eRNAs function in enhancer-directed transcriptional regulation, as well as if these elements are necessary for enhancer function. Future insights into the functional aspects of eRNAs will serve to clarify these discrepancies.

Clearly, the presence of transcription at enhancer loci can serve as a genomic signature of an active enhancer. It should be noted that in most of the studies examined, not all enhancers exhibit local transcription. In fact, some studies have suggested that eRNA production may be a coincidence of spurious transcription by RNAPII at enhancer-promoter proximal sites [51]. However, there has been growing evidence pointing to the contrary [52,53]. In the cases where eRNA transcripts were identified, most, if not all of their cognate genomic sequence were verified as active enhancer loci. The field of eRNAs remains a controversial and hotly debated topic, with new evidence emerging almost daily. However, the fact remains that these elements can be used to positively identify the locations of enhancers genome-wide. At the time of this review, no systematic evaluation of plant eRNAs has been conducted, resembling a lapse in our understanding of these mysterious elements. Nevertheless, eRNAs provide another potential mark for prediction and identification of plant enhancers.

3. Enhancer trapping in plants

Several enhancers such as the *tb1*, *b1*, *Bx1*, *Vgt1* and *KRN4* loci in Maize were some of the first plant enhancers identified, specifically through genetic mapping, though this technique is not a general approach or specifically designed for enhancer identification [54–58]. Enhancer traps represent the first general methodology developed for direct identification and functional validation of enhancer activity. The first enhancer traps were initially developed in *Drosophila* using the *lacZ* reporter fused to a weak promoter [59]. Early plant enhancer trapping methods were developed based on T-DNA insertion lines carrying promoter-less antibiotic resistance genes [60,61]. Newer generation enhancer traps contain a minimal promoter including a TATA box and a transcription start site, which is fused to the 5' end of a reporter gene, typically β -galactosidase (GUS) or green fluorescent protein (GFP). The minimal promoter is unable to drive transcription of the reporter gene unless in the presence of a transcription activating CRE [62]. Thus, identification of enhancers through this method relies on the insertion proximity of the T-DNA construct to a putative enhancer.

Enhancer traps have proven useful for enhancer identification in plants. Using this approach Fridborg et al. (2004) identified enhancers responsible for pathogen mediated responses in *Arabidopsis* [63]. Transformed *Arabidopsis* plants carrying the GUS reporter gene fused to a *Cauliflower mosaic virus* 35S minimal promoter displayed strong GUS signals when subjected to Tobacco Rattle Virus (TRV) [63]. Two enhancers located 1.1 kb away from the T-DNA insertion were identified and cloned, in addition to their target gene, a WRKY transcription factor, termed TRV-induced gene (*TRI*). Enhancer traps have been implemented in several plant species, including *Arabidopsis thaliana* [64,65], rice (*Oryza sativa*) [66,67], poplar (*Populus trichocarpa*) [68] and tobacco (*Nicotiana tabacum*) [69], owing to the functionality of these traps in heterologous species.

Unfortunately, enhancer traps have a few major limitations. Reporter gene activation is dependent on the proximity of the reporter construct to an active enhancer. Cellular expression of the reporter gene only indicates that a potential enhancer is near by the insertion site. Since enhancers can act in a distance and orientation independent manner, revealing the physical location of the enhancer with respect to the insertion is often a major challenge associated with trapping lines. Isolation of the enhancer sequence from a trapping line is often associated with tedious cloning and functional assays [70]. Furthermore, in order to simply tag a putative enhancer, enormous numbers of transgenic lines need to be generated which is a considerably laborious and time-consuming effort, and is only feasible in species where transformation protocols are well-established. Collectively, these drawbacks

reduce the throughput of enhancer trapping, a major bottleneck in the efforts to identify and predict enhancer loci genome-wide for many plant species. As a result, only a handful of enhancers have been isolated using this technique in plants [54,71–74].

4. Plant enhancer prediction and identification based on open chromatin signatures

The presence of sequences hypersensitive to DNase I cleavage, specific post-translational histone modifications, and nucleosome positioning and phasing, have all been suggested as distinct hallmarks of active or latent enhancers in mammalian systems [23,26,75–78]. Until recently, it has been unknown if these genomic and chromatin characteristics of enhancer held for plant species.

4.1. DHSs in plants

Recently, some of the first plant DNase-seq studies have been conducted in rice [79], *A. thaliana* [26,80,81] and *Brachypodium distachyon* [82]. The first set of plant DHSs were identified in rice by evaluating chromatin accessibility in both callus and seedling tissues [79]. Surprisingly, a staggering 42% and 45% of DHSs were located in intergenic regions from the callus and seedling libraries, respectively, providing the first genome-wide DHS-based observations of chromatin architecture in plants [79]. Shortly after, a similar approach was used to examine the genome-wide properties of open chromatin between flower and leaf tissues in *A. thaliana* [26]. In contrast to the discoveries in rice, out of all DHSs identified in *A. thaliana*, approximately 15% were located in intergenic regions, representing likely distal enhancers, or other active CREs. DHSs within 1000 bp upstream of genes account for 45% and 27% of the total DHSs in *Arabidopsis* and rice genomes, respectively [25]. These results suggest that plant species with compact genomes such as *Arabidopsis*, may contain fewer distal CREs than those with large complex genomes.

The DHS studies in rice and *Arabidopsis* also demonstrated that DHSs are depleted of bulk nucleosomes [26,79], similar to the experimental reports in mammalian species. Genes associated with flower-specific DHSs in *Arabidopsis* were enriched with functions related to flower, seed, and fruit development. In addition, flower-specific DHSs were highly correlated with the binding sites associated with two floral regulators, AP1 and SEP3. Protein-binding footprints as a consequence of TF binding can be potentially revealed based on the analysis of high-quality DNase-seq datasets [26]. These analyses were not exclusive to promoter DHSs, but however, included a broad set of DHSs within genomic regions including the 5' untranslated region (UTR), 3' UTR, exons, introns, and intergenic sequences. Given that enhancers can be distributed over many genomic features and are typically distant from their target genes, DHSs located outside of the promoter regions are attractive targets as putative enhancers.

4.2. DHS-based enhancer prediction and validation

Identifying putative enhancers can be achieved by sifting through genome-wide DHS datasets for regions classically associated with enhancers such as intergenic and intronic sites. However, for true validation of an enhancer, biochemical assays are needed to establish functional activity. We were able to identify and validate leaf- and flower-specific enhancers by utilizing a subset of intergenic DHSs in *A. thaliana* [83]. The sequence of each predicted enhancer was cloned into a GUS-based reporter vector. The constructs were then transformed into wild type *Arabidopsis* plants and assayed for GUS expression. A total of 14 predicted enhancers were selected for validation. Ten of the 14 enhancer candidates (71%) consistently generated GUS signals in different tissues with different signal intensities. Among the 14 predicted enhancers selected for validation, three were leaf-specific DHSs and three were flower-specific DHSs. All three leaf-specific enhancers generated

distinct GUS expression patterns exclusively localized in leaves, but not in flowers. Two of three flower-specific enhancers generated GUS signals in flowers. These results showed that DHS-based enhancer prediction is a promising avenue for enhancer identification and provides a general approach for enhancer prediction in any plant species with a sequenced genome. In addition, tissue-specific enhancers can be reliably predicted based on tissue-specific DHSs.

4.3. Additional marks for plant enhancer prediction

4.3.1. Histone modifications

Although genomic enhancers are depleted of nucleosomes, histone proteins neighboring enhancer sites generally carry characteristic post-translational modifications, especially H3K27ac, which can be exploited as a chromatin signature of enhancers [84]. In rice, 27% of the H3K27ac marks localized to intergenic regions and introns, while 30% of intergenic DHSs overlapped with H3K27ac modifications, suggesting that only 30% of intergenic DHSs may be predicted as putative enhancers [85]. This is in stark contrast with mammalian model organisms where the power to detect enhancers by H3K27ac and H3K4me1 appears to be much stronger than in plants. This is likely due to mixing of complex heterogeneous plant cell-types during DNase-seq and ChIP-seq library preparation, in contrast to mammalian species, where researchers have the luxury of homogenous cell lines. Interestingly, in whole rice seedlings, intergenic DHSs are enriched with H3K27me3 [79]. Elevated levels of H3K27me3 is suggestive of interactions with the Polycomb-complex, a transcriptional repressor involved with silencing a large number of genes, and is typically associated with tissue-specific developmental processes [86,87].

Predicted tissue-specific enhancers in *Arabidopsis* were examined for association with various histone modifications [83]. Putative leaf-specific enhancers were significantly more enriched with H3K27me3 in floral tissues than in leaf tissues. By contrast, putative leaf-specific enhancers were indeed more enriched with H3K27ac in leaf tissues than in floral tissues. These results suggest, as in mammalian systems, that active enhancers are associated with the presence of H3K27ac while poised enhancers are generalized by the presence of H3K27me3. Nevertheless, the association of H3K27me3 and H3K27ac with *Arabidopsis* intergenic enhancers were clearly not as striking as those reported in mammalian species. Thus, the value of histone modification data for predicting individual enhancer will be limited in plants, likely a result of heterogeneous datasets.

4.3.2. RNAPII binding sites

Notably, RNA polymerase II can be used in conjunction with DNase-seq to identify candidate enhancers [44,51,88]. In *Arabidopsis*, increased levels of RNAPII occupancy has been detected at DHSs within promoters, and is strongly correlated with gene expression [89]. We examined the RNAPII occupancy associated with intergenic DHSs ($n = 7515$) in *Arabidopsis*. DHSs were visibly enriched with RNAPII-binding relative to the flanking regions as well as a random control dataset (Fig. 3). Clearly, the presence of RNAPII-binding sites concomitant with DHSs at non-promoter regions provides robust evidence for enhancer prediction.

4.3.3. Transcription of genomic enhancers

Several recent studies in mammalian species indicated that some active enhancers are actively transcribed and are marked by the presence of noncoding RNA [44,90]. Thus the presence of eRNAs at an enhancer locus can serve as an additional feature for the identification of active enhancers.

Some 13,200 noncoding RNAs (ncRNAs) have been identified in the *Arabidopsis* genome [91]. Surprisingly, we found that only approximately 20% of intergenic DHSs (1991 out of 10,044) overlapped with a ncRNA transcript [83]. To ensure that this enrichment was not due to chance, we randomly selected a set of 10,044 intergenic sequences (300 bp

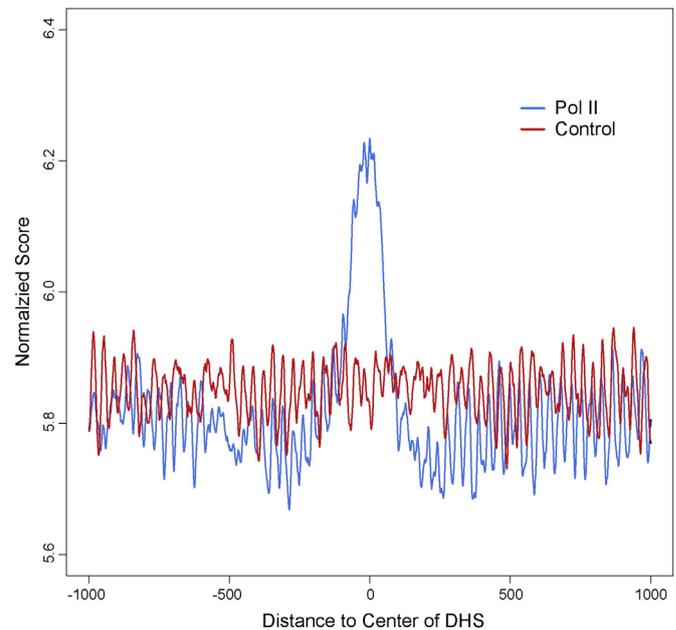


Fig. 3. RNA polymerase II localization at intergenic DHSs in *Arabidopsis*. DHSs were aligned by their midpoint coordinates and the normalized RNA Pol II scores were averaged using all intergenic DHSs ($n = 7515$). Random positions ($n = 7515$) without DHSs of similar length and composition as the DHS dataset was used as a control. The RNAP II signal rapidly decays as the distance from the DHS center increases, suggesting that intergenic DHSs are strongly associated with RNA Pol II.

long, the average length of the intergenic DHSs) and examined for the potential overlap of these sequences with ncRNAs, permuting this test 10,000 times. We found that the predicted enhancers were significantly more associated with overlap from ncRNAs than by chance [83]. However, the transcription levels of most mammalian eRNAs are low [44]. Thus, we predict a similarly low transcription level of plant eRNAs. It will be technically challenging to detect such low levels of transcription from tissue-specific plant enhancers because of the mixing of heterogeneous cell-types.

4.3.4. Nucleosome positioning and occupancy associated with DHSs

Nucleosome occupancy and positioning in the genome can be identified via chromatin digestion with *Micrococcal* nuclease (MNase) followed by sequencing (MNase-seq). Nucleosome-depleted regions (NDRs) are hypersensitive to DNase I digestion, and thus, are identified as DHSs with DNase-seq. Interestingly, we observed that intergenic DHSs in rice are often tightly associated with phased neighboring nucleosomes, a phenomenon typically observed at/or nearby actively transcribed genes [23,89]. It has also been reported that active promoters and regulatory sites are not in fact nucleosome depleted, but rather contain instable histone variants, H3.3 and H2A.Z [92]. Thus, information of nucleosome dynamics associated with DHSs, including details of nucleosome occupancy, the presence of H2A.Z, or flanking of phased nucleosomes, may provide additional evidence for enhancer prediction [89]. However, the correlations between DHSs and nucleosome dynamics were observed based on genome-wide analyses. The predicting power for a specific genomic region will rely on the quality of each genomic dataset, with careful attention directed to the impact of mixing cell-types associated with plant tissues during library construction.

5. Conclusions and future directions

Enhancers were traditionally difficult to identify because of their unpredictable positions. The advent of sequencing technologies and their applications in genome-wide chromatin studies have propelled enhancer research to new frontiers. Most enhancers can be predicted based on

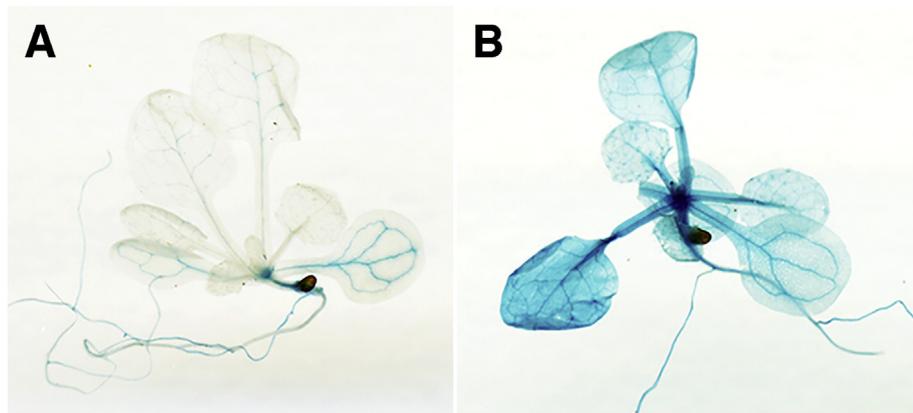


Fig. 4. Functional assay of a cold-induced intergenic enhancer candidate. The two transgenic plants (T_1) were derived from the same transgenic plant (T_0) carrying a cold-inducible DHS fused to the minimal 35S promoter neighboring the GUS reporter gene. Seeds were germinated at the same time point and were allowed to grow under standard conditions. (A) GUS assay of two-week-old seedling grown under room temperature. (B) GUS assay of a seedling that was placed under 4 °C for 24 h before the assay. Note: GUS signals were mainly observed in roots and main veins before cold treatment. GUS signals can be observed in the entire seedling after cold treatment.

their association with transcriptional activators and several unique features associated with open chromatin in mammalian species. Enhancer predicting methodologies can potentially be adapted for use in plants. However, the predicting power of genome-wide chromatin datasets in plants will be significantly affected by the heterogeneity of the data associated with mixing numerous cell-types in most plant tissues. Each data point from DNase-seq, ChIP-seq, or MNase-seq represents the average of data from multiple cell-types as well from cells at different developmental stages. This results in data which are considerably noisier than datasets generated by pure cell lines. Furthermore, since data generated in this manner represent the average of the sampled cell-types and developmental stages, there is a greater opportunity for type II error, or false negatives. True signal peaks from one specific cell-type can be lost or diminished due to the averaging with other non-signal bearing cell-types. This phenomenon can occur differentially between cells-types, ultimately leading to noisy and un-reliable data if many complex cell-types are used during the experimental procedure. Thus, development of datasets from individual cell-types will be a major factor to improve the accuracy of plant enhancer prediction based on chromatin features. Recently developed techniques, such as INTACT (isolation of nuclei tagged in specific cell-types) [93], could potentially be applied in order to circumvent these issues, allowing the development of cell type-specific chromatin datasets.

Traditional reporter-based assays provide the most direct functional evidence of predicted enhancers. We used the GUS reporter for enhancer validation [83]. The GUS assay is simple without the need for special equipment. However, it is destructive to transgenic plants, and thus cannot be used to detect enhancer activity throughout a plants life cycle. In addition, GUS assays are not an ideal method for quantitative measurement of enhancer function. A luciferase-based method would allow validation and quantitative measurements of enhancer activity throughout development stages of transgenic plants. Nevertheless, both GUS- and luciferase-based methods only allow validation of a single candidate enhancer at a time. The low-throughput nature of these methodologies represents a major bottleneck in attempts to characterize enhancers genome-wide. A recently developed technique, STARR-seq (self-transcribing active regulatory region sequencing), can be used to directly and quantitatively examine enhancer activity for millions of candidates from arbitrary sources of DNA [94]. This technique allows the detection, validation, and measurement of large number of enhancers in a single step. Additionally, CRISPR-Cas9 (clustered regularly interspaced short palindromic repeats) based gene editing method can be used to facilitate the identification of the gene(s) regulated by a targeted enhancer.

Plants have to respond and adapt to various environmental stresses and ultimately acquire stress tolerance for survival. Understanding

mechanisms for stress adaptation and tolerance is one of the most important and challenging goals in plant sciences and holds the key for future crop improvement. Identification of enhancers associated with genes related to stress responses will be one of the future frontiers in plant biology, as well as for plant breeding. Sullivan et al. (2014) developed DNase-seq datasets from heat-treated *Arabidopsis* seedlings and identified DHSs activated or repressed by heat shock [80]. We have recently conducted DNase-seq from cold-treated *Arabidopsis* seedlings. Cold-induced DHSs were readily identified from the DNase-seq dataset. We have also confirmed the enhancer activity of a cold-induced DHS using the GUS reporter gene assay (Fig. 4). Thus, the DHS-based approach provides a promising methodology to study enhancers associated with a variety of biotic and abiotic stresses in the future.

Transparency document

The Transparency document associated with this article can be found, in online version.

Acknowledgments

This was supported by grant MCB-1412948 from the National Science Foundation and a Vilas Associate Fellowship from the University of Wisconsin-Madison to J.J.

References

- [1] K. Kaufmann, A. Pajoro, G.C. Angenent, Regulation of transcription in plants: mechanisms controlling developmental switches, *Nat. Rev. Genet.* 11 (2010) 830–842.
- [2] E. Sparks, G. Wachsman, P.N. Benfey, Spatiotemporal signalling in plant development, *Nat. Rev. Genet.* 14 (2013) 631–644.
- [3] Y. Jiao, O.S. Lau, X.W. Deng, Light-regulated transcriptional networks in higher plants, *Nat Rev Genet* 8 (2007) 217–230.
- [4] I.K. Nolis, D.J. McKay, E. Mantouvalou, S. Lomvardas, M. Merika, D. Thanos, Transcription factors mediate long-range enhancer-promoter interactions, *Proc. Natl. Acad. Sci. U. S. A.* 106 (2009) 20222–20227.
- [5] J. Banerji, S. Rusconi, W. Schaffner, Expression of a β -globin gene is enhanced by remote SV40 DNA sequences, *Cell* 27 (1981) 299–308.
- [6] A. Williams, C.G. Spilianakis, R.A. Flavell, Interchromosomal association and gene regulation in trans, *Trends Genet.* 26 (2010) 188–197.
- [7] C.R. Clapier, B.R. Cairns, The biology of chromatin remodeling complexes, *Annu. Rev. Biochem.* 78 (2009) 273–304.
- [8] The ENCODE Project Consortium, An integrated encyclopedia of DNA elements in the human genome, *Nature* 489 (2012) 57–74.
- [9] M.G. Rosenfeld, V.V. Lunyak, C.K. Glass, Sensors and signals: a coactivator/corepressor/epigenetic code for integrating signal-dependent programs of transcriptional response, *Genes Dev.* 20 (2006) 1405–1428.
- [10] R.J. Kelleher, P.M. Flanagan, R.D. Kornberg, A novel mediator between activator proteins and the RNA polymerase II transcription apparatus, *Cell* 61 (1990) 1209–1215.

- [11] R.C. Conaway, S. Sato, C. Tomomori-Sato, T. Yao, J.W. Conaway, The mammalian Mediator complex and its role in transcriptional regulation, *Trends Biochem. Sci.* 30 (2005) 250–255.
- [12] M.H. Kuo, C.D. Allis, In vivo cross-linking and immunoprecipitation for studying dynamic Protein:DNA associations in a chromatin environment, *Methods* 19 (1999) 425–433.
- [13] K. Kaufmann, J.M. Muiño, M. Østerås, L. Farinelli, P. Krajewski, G.C. Angenent, Chromatin immunoprecipitation (ChIP) of plant transcription factors followed by sequencing (ChIP-SEQ) or hybridization to whole genome arrays (ChIP-CHIP), *Nat. Protoc.* 5 (2010) 457–472.
- [14] A. Visel, M.J. Blow, Z. Li, T. Zhang, J.A. Akiyama, A. Holt, I. Plajzer-Frick, M. Shoukry, C. Wright, F. Chen, V. Afzal, B. Ren, E.M. Rubin, L.A. Pennacchio, ChIP-seq accurately predicts tissue-specific activity of enhancers, *Nature* 457 (2009) 854–858.
- [15] M.H. Kagey, J.J. Newman, S. Bilodeau, Y. Zhan, D.A. Orlando, N.L. van Berkum, C.C. Ebmeier, J. Goossens, P.B. Rahl, S.S. Levine, D.J. Taatjes, J. Dekker, R.A. Young, Mediator and cohesin connect gene expression and chromatin architecture, *Nature* 467 (2010) 430–435.
- [16] J. Wang, J. Zhuang, S. Iyer, X. Lin, T.W. Whitfield, M.C. Greven, B.G. Pierce, X. Dong, A. Kundaje, Y. Cheng, O.J. Rando, E. Birney, R.M. Myers, W.S. Noble, M. Snyder, Z. Weng, Sequence features and chromatin structure around the genomic regions bound by 119 human transcription factors, *Genome Res.* 22 (2012) 1798–1812.
- [17] M. Tsompana, M.J. Buck, Chromatin accessibility: a window into the genome, *Epigenetics Chromatin* 7 (2014) 33.
- [18] J.M. Jiang, The 'dark matter' in the plant genomes: non-coding and unannotated DNA sequences associated with open chromatin, *Curr. Opin. Plant Biol.* 24 (2015) 17–23.
- [19] T.P. Chen, S.Y.R. Dent, Chromatin modifiers and remodellers: regulators of cellular differentiation, *Nat. Rev. Genet.* 15 (2014) 93–106.
- [20] C. Liu, D. Weigel, Chromatin in 3D: progress and prospects for plants, *Genome Biol.* 16 (2015) 170.
- [21] L.L. Wallrath, Q. Lu, H. Granok, S.C. Elgin, Architectural variations of inducible eukaryotic promoters: preset and remodeling chromatin structures, *BioEssays* 16 (1994) 165–170.
- [22] G.E. Crawford, I.E. Holt, J.C. Mullikin, D. Tai, R. Blakesley, G. Bouffard, A. Young, C. Masiello, E.D. Green, T.G. Wolfsberg, F.S. Collins, Identifying gene regulatory elements by genome-wide recovery of DNase hypersensitive sites, *Proc. Natl. Acad. Sci. U. S. A.* 101 (2004) 992–997.
- [23] Y.F. Wu, W.L. Zhang, J.M. Jiang, Genome-wide nucleosome positioning is orchestrated by genomic regions associated with DNase I hypersensitivity in rice, *PLoS Genet.* 10 (2014), e1004378.
- [24] W.L. Zhang, J.M. Jiang, Genome-wide mapping of DNase I hypersensitive sites in plants, *Methods Mol. Biol.* 1284 (2015) 71–89.
- [25] W.L. Zhang, T. Zhang, Y.F. Wu, J.M. Jiang, Open chromatin in plant genomes, *Cytogenetic and Genome Research* 143 (2014) 18–27.
- [26] W.L. Zhang, T. Zhang, Y.F. Wu, J.M. Jiang, Genome-wide identification of regulatory DNA elements and protein-binding footprints using signatures of open chromatin in *Arabidopsis*, *Plant Cell* 24 (2012) 2719–2731.
- [27] A.M. Sullivan, K.L. Bubb, R. Sandstrom, J.A. Stamatoyannopoulos, C. Queitsch, DNase I Hypersensitivity Mapping, Genomic Footprinting, and Transcription Factor Networks in Plants (Current Plant Biology) 2015.
- [28] R.E. Thurman, E. Rynes, R. Humbert, J. Vierstra, M.T. Maurano, E. Haugen, N.C. Sheffield, A.B. Stergachis, H. Wang, B. Vernot, K. Garg, S. John, R. Sandstrom, D. Bates, L. Boatman, T.K. Canfield, M. Diegel, D. Dunn, A.K. Ebersol, T. Frum, E. Giste, A.K. Johnson, E.M. Johnson, T. Kutuyavin, B. Lajoie, B.-K. Lee, K. Lee, D. London, D. Lotakis, S. Neph, F. Neri, E.D. Nguyen, H. Qu, A.P. Reynolds, V. Roach, A. Safi, M.E. Sanchez, A. Sanyal, A. Shafer, J.M. Simon, L. Song, S. Vong, M. Weaver, Y. Yan, Z. Zhang, Z. Zhang, B. Lenhard, M. Tewari, M.O. Dorschner, R.S. Hansen, P.A. Navas, G. Stamatoyannopoulos, V.R. Iyer, J.D. Lieb, S.R. Sunyaev, J.M. Akey, P.J. Sabo, R. Kaul, T.S. Furey, J. Dekker, G.E. Crawford, J.A. Stamatoyannopoulos, The accessible chromatin landscape of the human genome, *Nature* 489 (2012) 75–82.
- [29] M.S. Wilken, J.A. Brzezinski, A. La Torre, K. Siebenthal, R. Thurman, P. Sabo, R.S. Sandstrom, J. Vierstra, T.K. Canfield, R.S. Hansen, M.A. Bender, J. Stamatoyannopoulos, T.A. Reh, DNase I hypersensitivity analysis of the mouse brain and retina identifies region-specific regulatory elements, *Epigenetics Chromatin* 8 (2015) 8.
- [30] T. Kouzarides, Chromatin modifications and their function, *Cell* 128 (2007) 693–705.
- [31] E. Calo, J. Wysocka, Modification of enhancer chromatin: what, how, and why? *Mol. Cell* 49 (2013) 825–837.
- [32] S. Bonn, R.P. Zinzen, C. Girardot, E.H. Gustafson, A. Perez-Gonzalez, N. Delhomme, Y. Ghavi-Helm, B. Wilczynski, A. Riddell, E.E.M. Furlong, Tissue-specific analysis of chromatin state identifies temporal signatures of enhancer activity during embryonic development, *Nat. Genet.* 44 (2012) 148–156.
- [33] N.D. Heintzman, R.K. Stuart, G. Hon, Y. Fu, C.W. Ching, R.D. Hawkins, L.O. Barrera, S. Van Calcar, C. Qu, K.A. Ching, W. Wang, Z. Weng, R.D. Green, G.E. Crawford, B. Ren, Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome, *Nat. Genet.* 39 (2007) 311–318.
- [34] Y. Tian, Z. Jia, J. Wang, Z. Huang, J. Tang, Y. Zheng, Y. Tang, Q. Wang, Z. Tian, D. Yang, Y. Zhang, X. Fu, J. Song, S. Liu, J.C. van Velkinburgh, Y. Wu, B. Ni, Global mapping of H3K4me1 and H3K4me3 reveals the chromatin state-based cell type-specific gene regulation in human Treg cells, *PLoS One* 6 (2011), e27770.
- [35] M.P. Creyghton, A.W. Cheng, G.G. Welstead, T. Kooistra, B.W. Carey, E.J. Steine, J. Hanna, M.A. Lodato, G.M. Frampton, P.A. Sharp, L.A. Boyer, R.A. Young, R. Jaenisch, Histone H3K27ac separates active from poised enhancers and predicts developmental state, *Proc. Natl. Acad. Sci. U. S. A.* 107 (2010) 21931–21936.
- [36] G.E. Zentner, P.J. Tesar, P.C. Scacheri, Epigenetic signatures distinguish multiple classes of enhancers with distinct cellular functions, *Genome Res.* 21 (2011) 1273–1283.
- [37] N.L. Vastenhouw, A.F. Schier, Bivalent histone modifications in early embryogenesis, *Curr. Opin. Cell Biol.* 24 (2012) 374–386.
- [38] Y. Ghavi-Helm, F.A. Klein, T. Pakozdi, L. Ciglar, D. Noordermeer, W. Huber, E.E.M. Furlong, Enhancer loops appear stable during development and are associated with paused polymerase, *Nature* 512 (2014) 96–100.
- [39] L.E. DeMare, J. Leng, J. Cotney, S.K. Reilly, J. Yin, R. Sarro, J.P. Noonan, The genomic landscape of cohesin-associated chromatin interactions, *Genome Res.* 23 (2013) 1224–1234.
- [40] Y. Zhang, C.-H. Wong, R.Y. Birnbaum, G. Li, R. Favaro, C.Y. Ngan, J. Lim, E. Tai, H.M. Poh, E. Wong, F.H. Mulawadi, W.-K. Sung, S. Nicolis, N. Ahituv, Y. Ruan, C.-L. Wei, Chromatin connectivity maps reveal dynamic promoter-enhancer long-range associations, *Nature* 504 (2013) 306–310.
- [41] F. Jin, Y. Li, J.R. Dixon, S. Selvaraj, Z. Ye, A.Y. Lee, C.-A. Yen, A.D. Schmitt, C.A. Espinoza, B. Ren, A high-resolution map of the three-dimensional chromatin interactome in human cells, *Nature* 503 (2013) 290–294.
- [42] E. de Wit, W. de Laat, A decade of 3C technologies: insights into nuclear organization, *Genes Dev.* 26 (2012) 11–24.
- [43] D. Tuan, S.M. Kong, K. Hu, Transcription of the hypersensitive site HS2 enhancer in erythroid cells, *Proc. Natl. Acad. Sci. U. S. A.* 89 (1992) 11219–11223.
- [44] T.-K. Kim, M. Hemberg, J.M. Gray, A.M. Costa, D.M. Bear, J. Wu, D.A. Harmin, M. Laptewicz, K. Barbara-Haley, S. Kuersten, E. Markenscoff-Papadimitriou, D. Kuhl, H. Bitto, P.F. Worley, G. Kreiman, M.E. Greenberg, Widespread transcription at neuronal activity-regulated enhancers, *Nature* 465 (2010) 182–187.
- [45] W. Li, D. Notani, Q. Ma, B. Tanasa, E. Nunez, A.Y. Chen, D. Merkurjev, J. Zhang, K. Ohgi, X. Song, S. Oh, H.-S. Kim, C.K. Glass, M.G. Rosenfeld, Functional roles of enhancer RNAs for oestrogen-dependent transcriptional activation, *Nature* 498 (2013) 516–520.
- [46] N. Hah, S. Murakami, A. Nagari, C.G. Danko, W.L. Kraus, Enhancer transcripts mark active estrogen receptor binding sites, *Genome Res.* 23 (2013) 1210–1223.
- [47] J.-H. Cheng, D.Z.-C. Pan, Z.T.-Y. Tsai, H.-K. Tsai, Genome-wide analysis of enhancer RNA in gene regulation across 12 mouse tissues, *Sci. Rep.* 5 (2015) 12648.
- [48] C.-L. Hsieh, T. Fei, Y. Chen, T. Li, Y. Gao, X. Wang, T. Sun, C.J. Sweeney, G.-S.M. Lee, S. Chen, S.P. Balk, X.S. Liu, M. Brown, P.W. Kantoff, Enhancer RNAs participate in androgen receptor-driven looping that selectively enhances gene activation, *Proc. Natl. Acad. Sci. U. S. A.* 111 (2014) 7319–7324.
- [49] K. Mousavi, H. Zare, S. Dell'orso, L. Grontved, G. Gutierrez-Cruz, A. Derfoul, G.L. Hager, V. Sartorelli, eRNAs promote transcription by establishing chromatin accessibility at defined genomic loci, *Mol. Cell* 51 (2013) 606–617.
- [50] M.T.Y. Lam, H. Cho, H.P. Lesch, D. Gosselin, S. Heinz, Y. Tanaka-Oishi, C. Benner, M.U. Kaikkonen, A.S. Kim, M. Kosaka, C.Y. Lee, A. Watt, T.R. Grossman, M.G. Rosenfeld, R.M. Evans, C.K. Glass, Rev-Erbs repress macrophage gene expression by inhibiting enhancer-directed transcription, *Nature* 498 (2013) 511–515.
- [51] F. De Santa, I. Barozzi, F. Mietton, S. Ghisletti, S. Polletti, B.K. Tusi, H. Muller, J. Ragoussis, C.L. Wei, G. Natoli, A large fraction of extragenic RNA Pol II transcription sites overlap enhancers, *PLoS Biol.* 8 (2010), e1000384.
- [52] K. Pulakanti, L. Pinello, C. Stelloh, S. Blinka, J. Allred, S. Milanovich, S. Kiblawi, J. Peterson, A. Wang, G.C. Yuan, S. Rao, Enhancer transcribed RNAs arise from hypomethylated, Tet-occupied genomic regions, *Epigenetics* 8 (2013) 1303–1320.
- [53] M.T.Y. Lam, W. Li, M.G. Rosenfeld, C.K. Glass, Enhancer RNAs and regulated transcriptional programs, *Trends Biochem. Sci.* 39 (2014) 170–182.
- [54] R.M. Clark, T.N. Wagler, P. Quijada, J. Doebley, A distant upstream enhancer at the maize domestication gene *tb1* has pleiotropic effects on floral and inflorescent architecture, *Nat. Genet.* 38 (2006) 594–597.
- [55] M. Stam, C. Bebele, J.E. Dorweiler, V.L. Chandler, Differential chromatin structure within a tandem array 100 kb upstream of the maize *b1* locus is associated with paramutation, *Genes Dev.* 16 (2002) 1906–1918.
- [56] L.L. Zheng, M.D. McMullen, E. Bauer, C.C. Schon, A. Gierl, M. Frey, 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 (2015) 3917–3930.
- [57] S. Salvi, G. Sponza, M. Morgante, D. Tomes, X. Niu, K.A. Fengler, R. Meeley, E.V. Ananiev, S. Svitashv, E. Bruggemann, B. Li, C.F. Hainey, S. Radovic, G. Zaina, J.A. Rafalski, S.V. Tingey, G.H. Miao, R.L. Phillips, R. Tuberosa, Conserved noncoding genomic sequences associated with a flowering-time quantitative trait locus in maize, *Proc. Natl. Acad. Sci. U. S. A.* 104 (2007) 11376–11381.
- [58] L. Liu, Y.F. Du, X.M. Shen, M.F. Li, W. Sun, J. Huang, Z.J. Liu, Y.S. Tao, Y.L. Zheng, J.B. Yan, Z.X. Zhang, *KRN4* Controls Quantitative Variation in Maize Kernel Row Number, *PLoS Genetics*, 11, 2015.
- [59] C.J. O'Kane, W.J. Gehring, Detection in situ of genomic regulatory elements in *Drosophila*, *Proc. Natl. Acad. Sci. U. S. A.* 84 (1987) 9123–9127.
- [60] L. Herman, A. Jacobs, M. Van Montagu, A. Depicker, Plant chromosome/marker gene fusion assay for study of normal and truncated T-DNA integration events, *Mol. Gen. Genet.* 224 (1990) 248–256.
- [61] C. Koncz, N. Martini, R. Mayerhofer, Z. Koncz-Kalman, H. Korber, G.P. Redej, J. Schell, High-frequency T-DNA-mediated gene tagging in plants, *Proc. Natl. Acad. Sci. U. S. A.* 86 (1989) 8467–8471.
- [62] P.S. Springer, Gene traps: tools for plant development and genomics, *Plant Cell* 12 (2000) 1007–1020.
- [63] I. Fridborg, A. Williams, A. Yang, S. MacFarlane, K. Coutts, S. Angell, Enhancer trapping identifies TRI, an *Arabidopsis* gene up-regulated by pathogen infection, *Mol. Plant-Microbe Interact.* 17 (2004) 1086–1094.
- [64] V. Sundaresan, P. Springer, T. Volpe, S. Howard, J.D.G. Jones, C. Dean, H. Ma, R. Martienssen, Patterns of gene action in plant development revealed by enhancer trap and gene trap transposable elements, *Genes Dev.* 9 (1995) 1797–1810.

- [65] L. Campisi, Y.Z. Yang, Y. Yi, E. Heilig, B. Herman, A.J. Cassista, D.W. Allen, H.J. Xiang, T. Jack, Generation of enhancer trap lines in *Arabidopsis* and characterization of expression patterns in the inflorescence, *Plant J.* 17 (1999) 699–707.
- [66] C.Y. Wu, X.J. Li, W.Y. Yuan, G.X. Chen, A. Kilian, J. Li, C.G. Xu, X.H. Li, D.X. Zhou, S.P. Wang, Q.F. Zhang, Development of enhancer trap lines for functional analysis of the rice genome, *Plant J.* 35 (2003) 418–427.
- [67] Y.Z. Yang, H. Peng, H.M. Huang, J.X. Wu, S.R. Ha, D.F. Huang, T.G. Lu, Large-scale production of enhancer trapping lines for rice functional genomics, *Plant Sci.* 167 (2004) 281–288.
- [68] A. Groover, J.R. Fontana, G. Dupper, C.P. Ma, R. Martienssen, S. Strauss, R. Meilan, Gene and enhancer trap tagging of vascular-expressed genes in poplar trees, *Plant Physiol.* 134 (2004) 1742–1751.
- [69] A. Beilmann, K. Albrecht, S. Schultze, G. Wanner, U.M. Pfitzner, Activation of a truncated Pr-1 promoter by endogenous enhancers in transgenic plants, *Plant Mol. Biol.* 18 (1992) 65–78.
- [70] W. Yang, R.A. Jefferson, E. Huttner, J.M. Moore, W.B. Gagliano, U. Grossniklaus, An egg apparatus-specific enhancer of *Arabidopsis*, identified by enhancer detection, *Plant Physiol.* 139 (2005) 1421–1432.
- [71] R.C. McGarry, B.G. Ayre, 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 (2008) 839–849.
- [72] B. Raatz, A. Eicker, G. Schmitz, E. Fuss, D. Muller, S. Rossmann, K. Theres, Specific expression of LATERAL SUPPRESSOR is controlled by an evolutionarily conserved 3' enhancer, *Plant J.* 68 (2011) 400–412.
- [73] S.E. Schauer, P.M. Schluter, R. Baskar, J. Gheyselinck, A. Bolanos, M.D. Curtis, U. Grossniklaus, Intronic regulatory elements determine the divergent expression patterns of AGAMOUS-LIKE6 subfamily members in *Arabidopsis*, *Plant J.* 59 (2009) 987–1000.
- [74] W. Yang, R.A. Jefferson, E. Huttner, J.M. Moore, W.B. Gagliano, U. Grossniklaus, An egg apparatus-specific enhancer of *Arabidopsis*, identified by enhancer detection, *Plant Physiol.* 139 (2005) 1421–1432.
- [75] N. Harmston, B. Lenhard, Chromatin and epigenetic features of long-range gene regulation, *Nucleic Acids Res.* 41 (2013) 7185–7199.
- [76] C. Luo, J. Dong, Y. Zhang, E. Lam, Decoding the role of chromatin architecture in development: coming closer to the end of the tunnel, *Front. Plant Sci.* 5 (2014) 374.
- [77] A.J. Ruthenburg, H. Li, D.J. Patel, C.D. Allis, Multivalent engagement of chromatin modifications by linked binding modules, *Nat. Rev. Mol. Cell Biol.* 8 (2007) 983–994.
- [78] C. Wu, Y.-C. Wong, S.C.R. Elgin, The chromatin structure of specific genes: II. Disruption of chromatin structure during gene activity, *Cell* 16 (1979) 807–814.
- [79] W.L. Zhang, Y.F. Wu, J.C. Schnable, Z.X. Zeng, M. Freeling, G.E. Crawford, J.M. Jiang, High-resolution mapping of open chromatin in the rice genome, *Genome Res.* 22 (2012) 151–162.
- [80] A.M. Sullivan, A.A. Arsovski, J. Lempe, K.L. Bubb, M.T. Weirauch, P.J. Sabo, R. Sandstrom, R.E. Thurman, S. Neph, A.P. Reynolds, A.B. Stergachis, B. Vernot, A.K. Johnson, E. Haugen, S.T. Sullivan, A. Thompson, F.V. Neri, M. Weaver, M. Diegel, S. Mnaimneh, A. Yang, T.R. Hughes, J.L. Nemhauser, C. Queitsch, J.A. Stamatoyannopoulos, Mapping and dynamics of regulatory DNA and transcription factor networks in *A. thaliana*, *Cell Rep.* 8 (2014) 2015–2030.
- [81] J.S. Cumbie, S.A. Filichkin, M. Megraw, Improved DNase-seq protocol facilitates high resolution mapping of DNase I hypersensitive sites in roots in *Arabidopsis thaliana*, *Plant Methods* 11 (2015) 42.
- [82] T. Zhang, A.P. Marand, J.M. Jiang, PlantDHS: a database for DNase I hypersensitive sites in plants, *Nucleic Acids Res.* 44 (2016) D1148–D1153.
- [83] B. Zhu, W.L. Zhang, T. Zhang, B. Liu, J.M. Jiang, Genome-wide prediction and validation of intergenic enhancers in *Arabidopsis* using open chromatin signatures, *Plant Cell* 27 (2015) 2415–2426.
- [84] D. Heintzman, G.C. Hon, R.D. Hawkins, P. Kheradpour, A. Stark, L.F. Harp, Z. Ye, L.K. Lee, R.K. Stuart, C.W. Ching, K.A. Ching, J.E. Antosiewicz-Bourget, H. Liu, X. Zhang, R.D. Green, V.V. Lobanenkov, R. Stewart, J.A. Thomson, G.E. Crawford, M. Kellis, B. Ren, Histone modifications at human enhancers reflect global cell-type-specific gene expression, *Nature* 459 (2009) 108–112.
- [85] Z. Du, H. Li, Q. Wei, X. Zhao, C. Wang, Q. Zhu, X. Yi, W. Xu, X.S. Liu, W. Jin, Z. Su, Genome-wide analysis of histone modifications: H3K4me2, H3K4me3, H3K9ac, and H3K27ac in *Oryza sativa* L. *Japonica*, *Molecular Plant* 6 (2013) 1463–1472.
- [86] R. Andersson, C. Gebhard, I. Miguel-Escalada, I. Hoof, J. Bornholdt, M. Boyd, Y. Chen, X. Zhao, C. Schmidt, T. Suzuki, E. Ntini, E. Arner, E. Valen, K. Li, L. Schwarzfischer, D. Glatz, J. Raithe, B. Lilje, N. Rapin, F.O. Bagger, M. Jørgensen, P.R. Andersen, N. Bertin, O. Rackham, A.M. Burroughs, J.K. Baillie, Y. Ishizu, Y. Shimizu, E. Furuhashi, S. Maeda, Y. Negishi, C.J. Mungall, T.F. Meehan, T. Lassmann, M. Itoh, H. Kawaji, N. Kondo, J. Kawai, A. Lennartsson, C.O. Daub, P. Heutink, D.A. Hume, T.H. Jensen, H. Suzuki, Y. Hayashizaki, F. Müller, A.R.R. Forrest, P. Carninci, M. Rehli, A. Sandelin, An atlas of active enhancers across human cell types and tissues, *Nature* 507 (2014) 455–461.
- [87] X. Zhang, O. Clarenz, S. Cokus, Y.V. Bernatavichute, M. Pellegrini, J. Goodrich, S.E. Jacobsen, Whole-genome analysis of histone H3 lysine 27 trimethylation in *Arabidopsis*, *PLoS Biol.* 5 (2007), e129.
- [88] L. Song, Z.Z. Zhang, L.L. Grasfeder, A.P. Boyle, P.G. Giresi, B.-K. Lee, N.C. Sheffield, S. Gräf, M. Huss, D. Keefe, Z. Liu, D. London, R.M. McDaniel, Y. Shibata, K.A. Showers, J.M. Simon, T. Vales, T. Wang, D. Winter, Z.Z. Zhang, N.D. Clarke, E. Birney, V.R. Iyer, G.E. Crawford, J.D. Lieb, T.S. Furey, Open chromatin defined by DNaseI and FAIRE identifies regulatory elements that shape cell-type identity, *Genome Res.* 21 (2011) 1757–1767.
- [89] T. Zhang, W.L. Zhang, J.M. Jiang, Genome-wide nucleosome occupancy and positioning and their impact on gene expression and evolution in plants, *Plant Physiol.* 168 (2015) 1406–1416.
- [90] F. De Santa, I. Barozzi, F. Mietton, S. Ghisletti, S. Polletti, B.K. Tusi, H. Muller, J. Ragoussis, C.-L. Wei, G. Natoli, A large fraction of extragenic RNA pol II transcription sites overlap enhancers, *PLoS Biol.* 8 (2010), e1000384.
- [91] J. Liu, C. Jung, J. Xu, H. Wang, S. Deng, L. Bernad, C. Arenas-Huertero, N.-H. Chua, Genome-wide analysis uncovers regulation of long intergenic noncoding RNAs in *Arabidopsis*, *Plant Cell* 24 (2012) 4333–4345.
- [92] C. Jin, C. Zang, G. Wei, K. Cui, W. Peng, K. Zhao, G. Felsenfeld, H3.3/H2A.Z double variant-containing nucleosomes mark 'nucleosome-free regions' of active promoters and other regulatory regions, *Nat. Genet.* 41 (2009) 941–945.
- [93] R.B. Deal, S. Henikoff, The INTACT method for cell type-specific gene expression and chromatin profiling in *Arabidopsis thaliana*, *Nat. Protoc.* 6 (2011) 56–68.
- [94] C.D. Arnold, D. Gerlach, C. Stelzer, I.M. Boryn, M. Rath, A. Stark, Genome-wide quantitative enhancer activity maps identified by STARR-seq, *Science* 339 (2013) 1074–1077.