Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Plant Biology

Frontiers and techniques in plant gene regulation

Tobias Jores, Morgan Hamm, Josh T. Cuperus and Christine Queitsch

Abstract

Understanding plant gene regulation has been a priority for generations of plant scientists. However, due to its complex nature, the regulatory code governing plant gene expression has yet to be deciphered comprehensively. Recently developed methods—often relying on next-generation sequencing technology and state-of-the-art computational approaches—have started to further our understanding of the gene regulatory logic used by plants. In this review, we discuss these methods and the insights into the regulatory code of plants that they can yield.

Addresses

Department of Genome Sciences, University of Washington, Seattle, WA, USA

Corresponding authors: Jores, Tobias (tobjores@uw.edu); Cuperus, Josh T. (cuperusj@uw.edu); Queitsch, Christine (queitsch@uw.edu)

Current Opinion in Plant Biology 2023, **75:**102403

This review comes from a themed issue on **Epigenetics and gene regulation 2023**

Edited by **Nathan Springer** and **Hidetoshi Saze**

For complete overview of the section, please refer the article collection - Epigenetics and gene regulation 2023

Available online 16 June 2023

https://doi.org/10.1016/j.pbi.2023.102403

Keywords

Plant gene regulation, *Cis*-regulatory elements, DNA accessibility, Massively parallel reporter assay, Plant STARR-seq, Single-molecule chromatin assay, Single-cell sequencing.

Introduction

Faced with accelerating climate change and rapid population growth, we need crops with higher yields and greater resilience to ensure food security. Crop genome engineering will likely play an important role in generating these traits, and promising targets for such engineering are gene regulatory regions [1–4]. However, to enable targeted manipulation of plant gene expression, we need to better understand the plant regulatory code; specifically, we need to characterize plant regulatory elements and their interactions with each other and with *trans*-acting proteins (Figure 1).

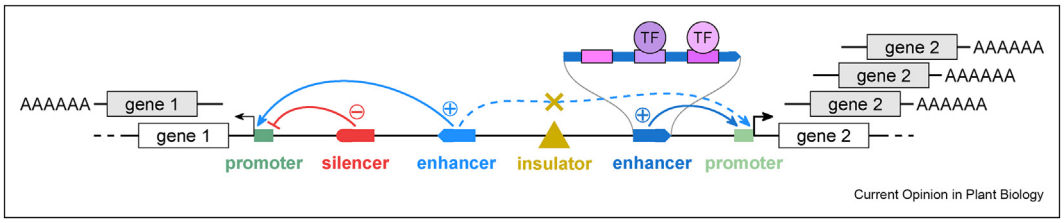
While the genetic code—mapping nucleotide triplets to amino acids—was elucidated decades ago, the regulatory code—linking regulatory DNA motifs to gene expression levels—still awaits comprehensive characterization. Three main differences between the genetic and the regulatory code make deciphering the latter a formidable challenge [5]: First, in contrast to the qualitative nature of the genetic code (which amino acid inserts at a protein position), the regulatory code has a strong quantitative component (how much a gene is expressed). Second, the genetic code consists of small, clearly defined and independent building blocks (*i.e.* codons), whereas the *cis*-regulatory elements that make up the regulatory code are larger, vary in size and are embedded in complex interaction networks. Third, the genetic code is largely universal (*i.e.* the same in almost all organisms, cell types and conditions), whereas the regulatory code is highly context-dependent and changes in response to environmental or developmental cues.

Despite the complexity of the regulatory code, great progress has been made towards its characterization [6,7]. Here, we review novel approaches in gene regulation research, and the aspects of the regulatory code that can be revealed with them.

Discovery of *cis*-regulatory elements

To decipher the regulatory code, we need to identify its building blocks. Gene expression is controlled by *cis*-regulatory elements such as core promoters, enhancers, silencers and insulators [6,7] (Figure 1). Core promoters reside at the transcription start site and contain DNA sequence elements that recruit the basal transcription machinery. Core promoters alone generally lead to low levels of gene expression, which increase through their interaction with enhancers [8]. Enhancers are DNA elements that serve as binding sites for condition- or tissue-specific transcription factors, which interact with the basal transcription machinery to drive transcription. Enhancers function independently of orientation, can reside upstream or downstream of their target core promoter, and are active over a wide range of distances [9,10]. Silencers share this orientation- and position-independency with enhancers; however, they decrease

Figure 1



Gene expression is controlled by the complex interaction network between *cis*-regulatory elements and *trans*-acting proteins. Core promoters recruit the core transcription machinery and establish a basal level of transcription. This expression level can be increased or decreased by enhancers and silencers, respectively. Insulators can block interactions among surrounding *cis*-regulatory elements. *Cis*-regulatory elements act by recruiting *trans*-acting proteins, like transcription factors (TFs), to mediate their function. Most elements contain multiple protein binding sites. Site occupancy depends on, for example, cell type and environmental conditions, leading to quantitative changes of gene expression.

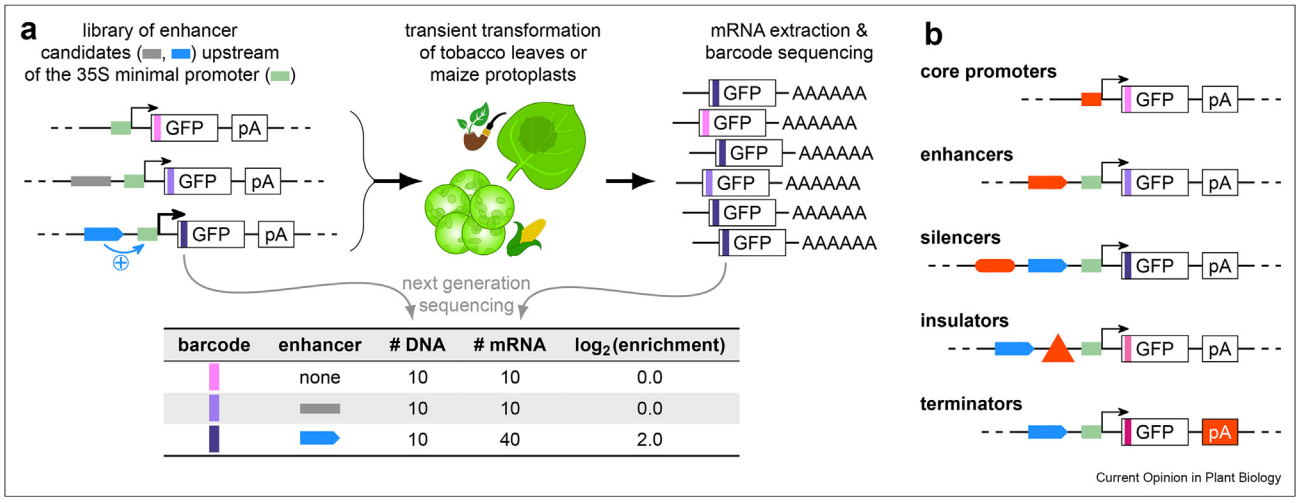
the expression of their target genes by recruiting repressive transcription factors. Both enhancers and silencers have the potential to interact with non-target genes; however, in the genome, such mis-regulation is prevented by DNA elements termed insulators, which block the action of distal enhancers or act as barriers that prevent the spread of condensed chromatin [11].

Until recently, the in-depth characterization of plant *cis*-regulatory elements relied on low-throughput techniques such as promoter deletion analysis, enhancer trapping and quantitative trait locus mapping. In 2013, a method was developed in *Drosophila* for the genome-wide identification of functional enhancers called self-transcribing active regulatory region sequencing (STARR-seq) [12]. STARR-seq inserts candidate enhancer sequences into the 3′ untranslated region of a reporter gene whose expression is under the control of a minimal promoter. If an insert has enhancer activity, it can upregulate its own transcription. The resulting transcript can be detected by next-generation sequencing and linked to its corresponding enhancer element incorporated within the mRNA. This strategy was used to identify enhancers in rice and maize protoplasts [10,13]. However, by carefully testing various configurations of the reporter construct, Jores et al. show that plant enhancers lose activity when placed in the 3′ untranslated region of the reporter gene. In contrast, when these enhancers are inserted immediately upstream of a core promoter, strong and condition-specific enhancer activity can be measured and enhancers can be characterized at single-nucleotide resolution [9]. This observation led to an optimized Plant STARR-seq assay with a large dynamic range and an improved signal-to-noise ratio (Figure 2a).

Moreover, because the identity of the regulatory elements to be tested is encoded by a DNA barcode, Plant STARR-seq can be adapted for promoters, insulators, silencers, untranslated regions and terminators by simply altering the position of the candidate elements within the reporter construct (Figure 2b). For example, Plant STARR-seq measured the strength of over 75,000 core promoters from *Arabidopsis*, sorghum and maize [8]. This study highlights the high-throughput nature of the assay, shows that it detects species- and condition-specific activity, and demonstrates its potential to enable the design of synthetic promoters. The approaches laid out in the promoter study can be followed to generate synthetic designs for other regulatory elements.

Although massively parallel reporter assays like Plant STARR-seq can rapidly expand our knowledge of plant regulatory elements, these assays rely on transient expression of a reporter construct and, therefore, lack the genomic context surrounding the tested elements. Systematic comparisons between results from such assays in transgenic plants, *Agrobacterium*-transformed plant cells and plasmid-transfected protoplasts have not been reported in plants. In human and *Drosophila* cells, studies could show strong concordance between the results obtained with plasmid-based and genome-integrated assays [12,14]. Since efficient systems for the position-specific integration of reporter constructs into plant genomes are lacking, other approaches have been used to study plant *cis*-regulatory elements in genomic context. For example, plant gene regulation at specific loci has been explored *in planta* using CRISPR/Cas9-mediated promoter bashing [2,15–18]. Here, CRISPR/Cas9 genome editing creates short deletions in the promoter region of target genes, and the edited plants are phenotyped to evaluate the impact of the deletion alleles. While this approach enables researchers to analyze *cis*-regulatory elements in their native genomic context, it is laborious and offers a far lower resolution than massively parallel assays. However, considered together, the results of massively parallel reporter assays can inform promoter bashing efforts to focus on the most promising genomic elements.

Figure 2



Plant STARR-seq: a massively parallel assay to study *cis*-regulatory elements in plants. (**a**) Enhancer candidates are cloned directly upstream of a 35S minimal promoter. The candidate sequences are linked to a short barcode located in the GFP reporter gene. The pooled library of reporter constructs is used for transient transformation of tobacco leaves or maize protoplasts. After 1–2 days incubation, the reporter mRNA is extracted. Finally, the relative abundance of the barcode sequences in the input DNA and in the output RNA is determined by next generation sequencing. Since enhancers drive increased transcription, the enrichment of a linked barcode in the RNA relative to its DNA input is a measure of enhancer strength. (**b**) Plant STARR-seq can be adapted to measure the activity of various different *cis*-regulatory elements.

Towards a mechanistic understanding of regulatory DNA elements

Cis-regulatory elements act by recruiting *trans*-acting proteins like transcription factors, which vary between different cell types and conditions. A mechanistic understanding of how *trans*-acting proteins interact with *cis*-regulatory elements is crucial if we want to predict cell-type- and condition-specific regulatory activity. For example, the DNA-binding preferences of plant transcription factors have been studied [19], with these efforts defining the DNA-binding preferences of over 600 plant transcription factors [20–25]. Recently, local DNA structure—often termed DNA shape—was shown to influence transcription factor binding. DNA shape can be used to improve the accuracy of transcription factor binding site predictions [26,27].

Despite the progress made towards predicting transcription factor binding to DNA, inferring how these binding events affect expression is difficult. Plant transcription factors often share similar DNA-binding preferences but act oppositely to either activate or repress nearby genes, depending on their—often unknown—effector domains. Although high-throughput screens for transcription activation or repression domains have been performed in yeast, *Drosophila* and humans [28–31], such systematic analyses of transcription factor effector domains have not yet been performed in plants. Using transcription factor binding predictions to infer gene regulation is further complicated by the fact that transcription factors can bind to DNA in a cooperative manner with altered binding site preferences and/or transcriptional output [24,32–34].

Due to the complexity of transcription factor-DNA interactions, several complementary approaches will likely be required to understand them. Both transcription factor-centric approaches and massively parallel reporter assays, like Plant STARR-seq, can lead to the identification of rules that govern transcription factor-DNA interactions. Analysis of how the order, spacing and orientation of transcription factor binding sites—the enhancer grammar [35]—affects the activity of regulatory DNA will help characterize additive and cooperative transcription factor interactions. Furthermore, massively parallel reporter assays could be used to directly measure the effect of one or more co-expressed transcription factor(s) on the activity of thousands of regulatory target sequences. Such experiments could also be coupled with the TARGET system to distinguish between direct and indirect effects of a transcription factor on gene expression [36,37].

Studying regulatory DNA interactions

Cis-regulatory elements do not work in isolation; they interact to accurately control gene expression. For example, while interactions between promoters and terminators affect gene expression, the relative strength of a terminator depends on which promoter it is paired

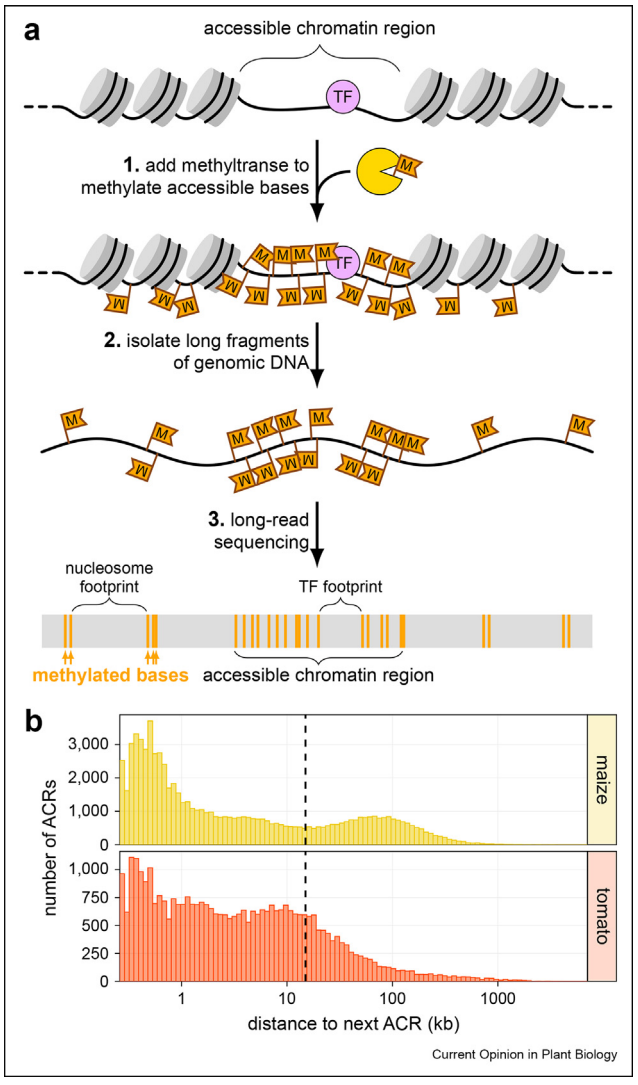
with, and *vice versa* [38]. Thus far, interactions between *cis*-regulatory elements remain understudied in plants. In flies, massively parallel reporter assays found that housekeeping and developmental promoters preferentially interact with enhancers of the same type [39,40]. Recent studies in human and mouse cells did not reach a consensus on whether promoter-enhancer interactions are promiscuous or specific [41,42]. Plant STARR-seq could be adapted to study not only interactions between plant promoters and enhancers, but also between other *cis*-regulatory elements. Learning the rules that govern regulatory DNA interactions will be crucial if we want to accurately predict and manipulate plant gene expression.

While massively parallel assays can be used to test the compatibility of *cis*-regulatory elements, identifying which elements interact in the context of the genome requires novel approaches to complement existing chromatin conformation data [6]. Potential regulatory interactions can be inferred from chromatin accessibility data by identifying co-accessible sites, *i.e.* sites with chromatin states that are correlated across samples or cells. Co-accessible sites can be identified from single-cell data [43,44]. However, due to the inherent sparsity of single-cell chromatin accessibility data (see below), accessibility patterns for cells belonging to a single cell type are typically aggregated. This aggregation masks heterogeneity at the level of individual chromatin molecules.

To overcome this limitation, single-molecule chromatin accessibility assays use long-read sequencing to probe chromatin state along individual chromatin fibers [45–48]. These assays use adenosine or cytosine methyltransferases to selectively methylate DNA in accessible regions. Long-read sequencing is then used to identify methylation-sensitive—and therefore accessible—regions along individual chromatin fibers (Figure 3a). These single-molecule assays capture the chromatin state of regulatory elements that can be separated by several kilobases or even megabases. By using long reads spanning two or more regulatory regions, the co-accessibility of these regions, in particular across changing conditions or cell types, can be determined and used to infer regulatory interactions.

Single-molecule chromatin assays can be readily adapted to plants, even to large crop genomes (Figure 3b). However, protocols for efficient use of single-molecule chromatin assays in plants have yet to be reported. Several technical hurdles need to be overcome for their successful application. For example, such assays require large amounts of nuclear input DNA, while contamination with mitochondrial or chloroplast DNA reduces the number of productive reads. Beyond co-accessibility analysis, these assays offer a variety of advantages over short-read-based methods to determine chromatin accessibility. Single-molecule chromatin assays also capture native CpG methylation, nucleosome positions, and footprints of DNA-binding proteins [45–48]. Further, these long-read-based assays allow the analysis of chromatin states in repetitive and transposon-rich regions of the genome which cannot be resolved by short reads. Such regions are particularly prevalent in

Figure 3



Assessing the chromatin state of individual DNA molecules. (**a**) Single-molecule chromatin assays use exogenously added methyltransferases to methylate accessible bases *in situ*. Long-read sequencing is then used to detect methylated bases. Accessible chromatin regions, nucleosome placement and footprints of DNA-binding proteins can be inferred from the methylation patterns. (**b**) Most accessible chromatin regions (ACRs) are close enough to another ACR to co-occur with it on a single read enabling the detection of co-accessibility in plant genomes, including the large genomes of major crops. The dashed line indicates the average length of a PacBio HiFi read (≈15 kb). Oxford Nanopore reads can be even longer.

crop genomes. Finally, single-molecule chromatin assays can be adapted to localize DNA-binding proteins and identify correlated long-range binding events that occur on individual chromatin fibers [49].

Deciphering cell-type-specific gene regulation

Gene regulation is highly context-dependent, such that the activity of *cis*-regulatory elements can change depending on environmental conditions and cell identity [50]. Studying gene regulation at single-cell resolution has the potential to resolve complex regulatory interactions and identify targets for genetic manipulation. Single-cell RNA sequencing (scRNA-seq), which measures cellular transcriptomes, and single-cell assay of transposase-accessible chromatin sequencing (scATAC-seq), which measures chromatin accessibility, have been used to study cell-type-dependent gene regulation in several plant species [51,52]. These studies discovered cell-type-specific changes in gene expression and chromatin accessibility and identified potentially causative transcription factors [53–56]. Furthermore, these studies show that integration of the richer scRNA-seq data with the sparser scATAC-seq data facilitates cell type annotations and data interpretation.

While single-cell genomics provides rich information about cell-type-specific gene regulation, this technology comes with several limitations [57]. One key challenge is data sparsity. This is especially true for scATAC-seq data where the count matrix (counting cut sites per cell and accessible chromatin region) is essentially binary (*i.e.* containing only zeros and ones) with less than 10% non-zero entries [58]. Additionally, single-cell experiments often struggle to capture sufficient cells of low-abundance or more recalcitrant cell types. While these limitations can be overcome, at least in part, by sequencing more cells with higher coverage, this solution can quickly become cost-prohibitive. However, recently developed single-cell genomics methods promise to drastically reduce the cost per cell as compared to commonly used commercial solutions [59–61].

Computational approaches to study plant gene regulation

The type, scale and complexity of the data generated by massively parallel assays and single-cell or single-molecule sequencing experiments require advanced computational methods for analysis and interpretation. Deep learning using artificial neural networks has proven especially useful for this purpose. Convolutional neural networks trained on data from Plant STARR-seq accurately predicted the activity of novel sequences, identified the underlying functional DNA features and improved regulatory activity through *in silico* evolution [8,62]. While developing, training and interpreting deep learning models can be a daunting task for non-specialists, recent work has aimed to make this process more accessible [63–65].

Genomics data can also be used to train artificial neural networks. For example, sequence-based deep learning models have been developed to predict transcription factor binding sites [66,67] and chromatin accessibility [68,69] in plants. Ideally, we would like to predict gene expression levels directly from DNA sequences to enable crop breeding and engineering. However, despite recent progress in this field [70–72], further work is required to make the corresponding models more accurate and generalizable, especially in plants.

Conclusions

With the advent of next-generation sequencing technology, several methods for the high-throughput analysis of regulatory DNA have been developed: Massively parallel reporter assays drastically increase the number of *cis*-regulatory elements that can be tested in an experiment. Single-molecule chromatin assays allow for inferences on interactions of regulatory elements and *trans*-acting proteins binding them. Single-cell genomics enables the study of gene regulation across many cell types in parallel.

Most of these technologies were developed in non-plant systems; however, adapting them for use in plants promises outsized knowledge gains. Compared to animal and human genomes, crop genomes typically contain many more repeats and far larger transcription factor families and show higher ploidy, all of which likely contributes to their environmental resilience and their capacity for continuous development. Combined, these technologies, advanced computational analyses, and theoretical studies will ultimately enable us to understand how gene regulation allows for the sessile life style of plants and for their exquisitely tuned development in concert with their biotic and abiotic environment. This thorough understanding will enable the successful engineering of stress-resistant and high-yielding crops of the future.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

Acknowledgments

This work was supported by the National Science Foundation (RESEARCH-PGR grant 1748843 and PlantSynBio 2240888 to C.Q.), and the German Research Foundation (DFG; fellowship 441540116 to T.J.).

References

Papers of particular interest, published within the period of review, have been highlighted as:

* of special interest
** of outstanding interest

1. Scheben A, Wolter F, Batley J, Puchta H, Edwards D: **Towards CRISPR/Cas crops – bringing together genomics and genome editing**. *New Phytol* 2017, **216**:682–698.

2. ** Wang X, Aguirre L, Rodríguez-Leal D, Hendelman A, Benoit M, Lippman ZB: **Dissecting cis-regulatory control of quantitative trait variation in a plant stem cell circuit**. *Nat Plants* 2021, **7**: 419–427.
Wang et al. utilized CRISPR/Cas9-mediated genome editing to create deletion alleles in the promoter of *CLV3* and important regulator of stem cell proliferation in tomato. The alleles lead to quantitative variation in the number of locules per fruit, demonstrating that fine-tuning of promoter strength can be used to modify yield-relevant traits.

3. Crisp PA, Bhatnagar-Mathur P, Hundleby P, Godwin ID, Waterhouse PM, Hickey LT: **Beyond the gene: epigenetic and cis-regulatory targets offer new breeding potential for the future**. *Curr Opin Biotechnol* 2022, **73**:88–94.

4. Zhong V, Archibald BN, Brophy JAN: **Transcriptional and post-transcriptional controls for tuning gene expression in plants**. *Curr Opin Plant Biol* 2023, **71**:102315.

5. Kim S, Wysocka J: **Deciphering the multi-scale, quantitative cis-regulatory code**. *Mol Cell* 2023, **83**:373–392.

6. Schmitz RJ, Grotewold E, Stam M: **Cis-regulatory sequences in plants: their importance, discovery, and future challenges**. *Plant Cell* 2022, **34**:718–741.

7. Marand AP, Eveland AL, Kaufmann K, Springer NM: ***cis*-Regulatory elements in plant development, adaptation, and evolution**. *Annu Rev Plant Biol* 2023:74. annurev-arplant-070122-030236.

8. ** Jores T, Tonnies J, Wrightsman T, Buckler ES, Cuperus JT, Fields S, Queitsch C: **Synthetic promoter designs enabled by a comprehensive analysis of plant core promoters**. *Nat Plants* 2021, **7**:842–855.
Characterization of over 75,000 core promoters from *Arabidopsis*, maize and sorghum using plant STARR-seq in tobacco leaves and maize protoplasts. This study identified key features of core promoters and developed a computational model that can predict promoter strength and improve it through *in silico* evolution.

9. Jores T, Tonnies J, Dorrity MW, Cuperus JT, Fields S, Queitsch C: **Identification of plant enhancers and their constituent elements by STARR-seq in tobacco leaves**. *Plant Cell* 2020, **32**:2120–2131.

10. Ricci WA, Lu Z, Ji L, Marand AP, Ethridge CL, Murphy NG, Noshay JM, Galli M, Mejía-Guerra MK, Colomé-Tatché M, *et al.*: **Widespread long-range cis-regulatory elements in the maize genome**. *Nat Plants* 2019, **5**:1237–1249.

11. Singer SD, Liu Z, Cox KD: **Minimizing the unpredictability of transgene expression in plants: the role of genetic insulators**. *Plant Cell Rep* 2012, **31**:13–25.

12. Arnold CD, Gerlach D, Stelzer C, Boryń ŁM, Rath M, Stark A: **Genome-wide quantitative enhancer activity maps identified by STARR-seq**. *Science* 2013, **339**:1074–1077.

13. Sun J, He N, Niu L, Huang Y, Shen W, Zhang Y, Li L, Hou C: **Global quantitative mapping of enhancers in rice by STARR-seq**. *Dev Reprod Biol* 2019, **17**:140–153.

14. Klein JC, Agarwal V, Inoue F, Keith A, Martin B, Kircher M, Ahituv N, Shendure J: **A systematic evaluation of the design and context dependencies of massively parallel reporter assays**. *Nat Methods* 2020, **17**:1083–1091.

15. Hendelman A, Zebell S, Rodriguez-Leal D, Dukler N, Robitaille G, Wu X, Kostyun J, Tal L, Wang P, Bartlett ME, *et al.*: **Conserved pleiotropy of an ancient plant homeobox gene uncovered by cis-regulatory dissection**. *Cell* 2021, **184**: 1724–1739.e16.

16. Liu L, Gallagher J, Arevalo ED, Chen R, Skopelitis T, Wu Q, Bartlett M, Jackson D: **Enhancing grain-yield-related traits by CRISPR–Cas9 promoter editing of maize CLE genes**. *Nat Plants* 2021, **7**:287–294.

17. Song X, Meng X, Guo H, Cheng Q, Jing Y, Chen M, Liu G, Wang B, Wang Y, Li J, *et al.*: **Targeting a gene regulatory element enhances rice grain yield by decoupling panicle number and size**. *Nat Biotechnol* 2022, **40**:1403–1411.

18. Rodríguez-Leal D, Lemmon ZH, Man J, Bartlett ME, Lippman ZB: **Engineering quantitative trait variation for crop improvement by genome editing**. *Cell* 2017, **171**:470–480.e8.

19. Lai X, Stigliani A, Vachon G, Carles C, Smaczniak C, Zubieta C, Kaufmann K, Parcy F: **Building transcription factor binding site models to understand gene regulation in plants**. *Mol Plant* 2019, **12**:743–763.

20. O'Malley RC, Huang SC, Song L, Lewsey MG, Bartlett A, Nery JR, Galli M, Gallavotti A, Ecker JR: **Cistrome and epicistrome features shape the regulatory DNA landscape**. *Cell* 2016, **165**:1280–1292.

21. Galli M, Khakhar A, Lu Z, Chen Z, Sen S, Joshi T, Nemhauser JL, Schmitz RJ, Gallavotti A: **The DNA binding landscape of the maize AUXIN RESPONSE FACTOR family**. *Nat Commun* 2018, **9**:4526.

22. Tian F, Yang D-C, Meng Y-Q, Jin J, Gao G: **PlantRegMap: charting functional regulatory maps in plants**. *Nucleic Acids Res* 2020, **48**:D1104–D1113.

23. Sun Y, Oh D-H, Duan L, Ramachandran P, Ramirez A, Bartlett A, Tran K-N, Wang G, Dassanayake M, Dinneny JR: **Divergence in the ABA gene regulatory network underlies differential growth control**. *Nat Plants* 2022, **8**:549–560.

24. Li M, Yao T, Lin W, Hinckley WE, Galli M, Muchero W, Gallavotti A, Chen J-G, Huang SC: **Double DAP-seq uncovered synergistic DNA binding of interacting bZIP transcription factors**. *Nat Commun* 2023, **14**:2600.

25. Tu X, Mejía-Guerra MK, Valdes Franco JA, Tzeng D, Chu P-Y, Shen W, Wei Y, Dai X, Li P, Buckler ES, *et al.*: **Reconstructing the maize leaf regulatory network using ChIP-seq data of 104 transcription factors**. *Nat Commun* 2020, **11**:5089.

26. * López-Vidriero I, Godoy M, Grau J, Peñuelas M, Solano R, Franco-Zorrilla JM: **DNA features beyond the transcription factor binding site specify target recognition by plant MYC2-related bHLH proteins**. *Plant Commun* 2021, **2**:100232.
Examination how features, including DNA shape, outside of the canonical core motif can contribute to and help to differentiate the binding specificities of paralogous transcription factors.

27. * Sielemann J, Wulf D, Schmidt R, Bräutigam A: **Local DNA shape is a general principle of transcription factor binding specificity in Arabidopsis thaliana**. *Nat Commun* 2021, **12**:6549.
Analysis how DNA shape affects the binding preferences of transcription factors. Using a machine learning approach that incorporates DNA shape information led to improved accuracy in binding site predictions for over 200 transcription factors from *Arabidopsis*.

28. Arnold CD, Nemčko F, Woodfin AR, Wienerroither S, Vlasova A, Schleiffer A, Pagani M, Rath M, Stark A: **A high-throughput method to identify trans-activation domains within transcription factor sequences**. *EMBO J* 2018, **37**, e98896.

29. Erijman A, Kozlowski L, Sohrabi-Jahromi S, Fishburn J, Warfield L, Schreiber J, Noble WS, Söding J, Hahn S: **A high-throughput screen for transcription activation domains reveals their sequence features and permits prediction by deep learning**. *Mol Cell* 2020, **78**:890–902.e6.

30. Tycko J, DelRosso N, Hess GT, Aradhana, Banerjee A, Mukund A, Van MV, Ego BK, Yao D, Spees K, *et al.*: **High-throughput discovery and characterization of human transcriptional effectors**. *Cell* 2020, **183**:2020–2035.e16.

31. Sanborn AL, Yeh BT, Feigerle JT, Hao CV, Townshend RJ, Lieberman Aiden E, Dror RO, Kornberg RD: **Simple biochemical features underlie transcriptional activation domain diversity and dynamic, fuzzy binding to Mediator**. *Elife* 2021, **10**, e68068.

32. Jolma A, Yin Y, Nitta KR, Dave K, Popov A, Taipale M, Enge M, Kivioja T, Morgunova E, Taipale J: **DNA-dependent formation of transcription factor pairs alters their binding specificity**. *Nature* 2015, **527**:384–388.

33. Ibarra IL, Hollmann NM, Klaus B, Augsten S, Velten B, Hennig J, Zaugg JB: **Mechanistic insights into transcription factor cooperativity and its impact on protein-phenotype interactions**. *Nat Commun* 2020, **11**:124.

34. ** van Mourik H, Chen P, Smaczniak C, Boeren S, Kaufmann K, Bemer M, Angenent GC, Muino JM: **Dual specificity and target gene selection by the MADS-domain protein FRUITFULL**. *Nat Plants* 2023:1–13.
This study uses both *in vitro* biochemistry and genomic techniques to analyze how tissue-specific, heteromeric complex formation of the transcription factor FRUITFULL alters its DNA binding preferences and leads to the regulation of different sets of target genes.

35. Jindal GA, Farley EK: **Enhancer grammar in development, evolution, and disease: dependencies and interplay**. *Dev Cell* 2021, **56**:575–587.

36. Alvarez JM, Schinke A-L, Brooks MD, Pasquino A, Leonelli L, Varala K, Safi A, Krouk G, Krapp A, Coruzzi GM: **Transient genome-wide interactions of the master transcription factor NLP7 initiate a rapid nitrogen-response cascade**. *Nat Commun* 2020, **11**:1157.

37. Brooks MD, Reed KM, Krouk G, Coruzzi GM, Bargmann BOR: **The TARGET system: rapid identification of direct targets of transcription factors by gene regulation in plant cells**. In *Transcription factor regulatory networks*. Edited by Song Q, Tao Z, *Methods in molecular biology*, **2594**. New York, NY: Humana; 2022:1–12.

38. Andreou AI, Nirkko J, Ochoa-Villarreal M, Nakayama N: **Mobius Assembly for Plant Systems highlights promoter-terminator interaction in gene regulation**. *bioRxiv* 2021. https://doi.org/10.1101/2021.03.31.437819.

39. Zabidi MA, Arnold CD, Schernhuber K, Pagani M, Rath M, Frank O, Stark A: **Enhancer–core-promoter specificity separates developmental and housekeeping gene regulation**. *Nature* 2015, **518**:556–559.

40. Arnold CD, Zabidi MA, Pagani M, Rath M, Schernhuber K, Kazmar T, Stark A: **Genome-wide assessment of sequence-intrinsic enhancer responsiveness at single-base-pair resolution**. *Nat Biotechnol* 2017, **35**:136–144.

41. Martinez-Ara M, Comoglio F, van Arensbergen J, van Steensel B: **Systematic analysis of intrinsic enhancer-promoter compatibility in the mouse genome**. *Mol Cell* 2022, **82**:2519–2531.e6.

42. Bergman DT, Jones TR, Liu V, Ray J, Jagoda E, Siraj L, Kang HY, Nasser J, Kane M, Rios A, *et al.*: **Compatibility rules of human enhancer and promoter sequences**. *Nature* 2022, **607**:176–184.

43. Pliner HA, Packer JS, McFaline-Figueroa JL, Cusanovich DA, Daza RM, Aghamirzaie D, Srivatsan S, Qiu X, Jackson D, Minkina A, *et al.*: **Cicero predicts cis-regulatory DNA interactions from single-cell chromatin accessibility data**. *Mol Cell* 2018, **71**:858–871.e8.

44. Gate RE, Cheng CS, Aiden AP, Siba A, Tabaka M, Lituiev D, Machol I, Gordon MG, Subramaniam M, Shamim M, *et al.*: **Genetic determinants of co-accessible chromatin regions in activated T cells across humans**. *Nat Genet* 2018, **50**:1140–1150.

45. Wang Y, Wang A, Liu Z, Thurman AL, Powers LS, Zou M, Zhao Y, Hefel A, Li Y, Zabner J, *et al.*: **Single-molecule long-read sequencing reveals the chromatin basis of gene expression**. *Genome Res* 2019, **29**:1329–1342.

46. Stergachis AB, Debo BM, Haugen E, Churchman LS, Stamatoyannopoulos JA: **Single-molecule regulatory architectures captured by chromatin fiber sequencing**. *Science* 2020, **368**:1449–1454.

47. Abdulhay NJ, McNally CP, Hsieh LJ, Kasinathan S, Keith A, Estes LS, Karimzadeh M, Underwood JG, Goodarzi H, Narlikar GJ, *et al.*: **Massively multiplex single-molecule oligonucleosome footprinting**. *Elife* 2020, **9**, e59404.

48. Lee I, Razaghi R, Gilpatrick T, Molnar M, Gershman A, Sadowski N, Sedlazeck FJ, Hansen KD, Simpson JT, Timp W: **Simultaneous profiling of chromatin accessibility and methylation on human cell lines with nanopore sequencing**. *Nat Methods* 2020, **17**:1191–1199.

49. Altemose N, Maslan A, Smith OK, Sundararajan K, Brown RR, Mishra R, Detweiler AM, Neff N, Miga KH, Straight AF, *et al.*: **DiMeLo-seq: a long-read, single-molecule method for mapping protein–DNA interactions genome wide**. *Nat Methods* 2022, **19**:711–723.

50. Sinha R, Zandalinas SI, Fichman Y, Sen S, Zeng S, Gómez-Cadenas A, Joshi T, Fritschi FB, Mittler R: **Differential regulation of flower transpiration during abiotic stress in annual plants**. *New Phytol* 2022, **235**:611–629.

51. Cuperus JT: **Single-cell genomics in plants: current state, future directions, and hurdles to overcome**. *Plant Physiol* 2022, **188**:749–755.

52. Marand AP, Schmitz RJ: **Single-cell analysis of cis-regulatory elements**. *Curr Opin Plant Biol* 2022, **65**:102094.

53. ** Dorrity MW, Alexandre CM, Hamm MO, Vigil A-L, Fields S, Queitsch C, Cuperus JT: **The regulatory landscape of Arabidopsis thaliana roots at single-cell resolution**. *Nat Commun* 2021, **12**:3334.
Study of plant gene regulation using combined single-cell RNA-seq and ATAC-seq data obtained from *Arabidopsis* roots. Co-embedding the RNA-seq and ATAC-seq data leads to improved annotation of cell-type and developmental stage.

54. ** Marand AP, Chen Z, Gallavotti A, Schmitz RJ: **A cis-regulatory atlas in maize at single-cell resolution**. *Cell* 2021, **184**:3041–3055.e21.
Large-scale interrogation of cell-type-specific gene regulation in maize using single-cell sequencing. Using accessibility information from over 50,000 cells, this study identified transcription factors that contribute to chromatin organization and co-accessibility.

55. Zhang T-Q, Chen Y, Liu Y, Lin W-H, Wang J-W: **Single-cell transcriptome atlas and chromatin accessibility landscape reveal differentiation trajectories in the rice root**. *Nat Commun* 2021, **12**:2053.

56. Farmer A, Thibivilliers S, Ryu KH, Schiefelbein J, Libault M: **Single-nucleus RNA and ATAC sequencing reveals the impact of chromatin accessibility on gene expression in Arabidopsis roots at the single-cell level**. *Mol Plant* 2021, **14**:372–383.

57. Lähnemann D, Köster J, Szczurek E, McCarthy DJ, Hicks SC, Robinson MD, Vallejos CA, Campbell KR, Beerenwinkel N, Mahfouz A, *et al.*: **Eleven grand challenges in single-cell data science**. *Genome Biol* 2020, **21**:31.

58. Li Z, Kuppe C, Ziegler S, Cheng M, Kabgani N, Menzel S, Zenke M, Kramann R, Costa IG: **Chromatin-accessibility estimation from single-cell ATAC-seq data with scOpen**. *Nat Commun* 2021, **12**:6386.

59. Martin BK, Qiu C, Nichols E, Phung M, Green-Gladden R, Srivatsan S, Blecher-Gonen R, Beliveau BJ, Trapnell C, Cao J, *et al.*: **Optimized single-nucleus transcriptional profiling by combinatorial indexing**. *Nat Protoc* 2023, **18**:188–207.

60. Datlinger P, Rendeiro AF, Boenke T, Senekowitsch M, Krausgruber T, Barreca D, Bock C: **Ultra-high-throughput single-cell RNA sequencing and perturbation screening with combinatorial fluidic indexing**. *Nat Methods* 2021, **18**:635–642.

61. Clark IC, Fontanez KM, Meltzer RH, Xue Y, Hayford C, May-Zhang A, D'Amato C, Osman A, Zhang JQ, Hettige P, *et al.*: Microfluidics-free single-cell genomics with templated emulsification. Nat Biotechnol 20231–10.

62. Deng K, Zhang Q, Hong Y, Yan J, Hu X: **iCREPCP: a deep learning-based web server for identifying base-resolution cis-regulatory elements within plant core promoters**. *Plant Commun* 2023, **4**:100455.

63. Chalupová E, Vaculík O, Poláček J, Jozefov F, Majtner T, Alexiou P: **ENNGene: an easy neural network model building tool for genomics**. *BMC Genom* 2022, **23**:248.

64. Klie A, Stites H, Jores T, Carter H, EUGENe: **A Python toolkit for predictive analyses of regulatory sequences**. *bioRxiv* 2022. https://doi.org/10.1101/2022.10.24.513593.

65. Dale R, Oswald S, Jalihal A, LaPorte M-F, Fletcher DM, Hubbard A, Shiu S-H, Nelson ADL, Bucksch A: **Overcoming the challenges to enhancing experimental plant biology with computational modeling**. *Front Plant Sci* 2021, **12**:687652.

66. Bang S, Galli M, Crisp PA, Gallavotti A, Schmitz RJ: **Identifying transcription factor–DNA interactions using machine learning**. *Silico Plants* 2022, **4**:diac014.

67. * Yan W, Li Z, Pian C, Wu Y: **PlantBind: an attention-based multi-label neural network for predicting plant transcription factor binding sites**. *Briefings Bioinf* 2022, **23**:bbac425.
Development of a deep learning model to predict the binding sites of over 300 transcription factors from *Arabidopsis* based on DNA sequence and shape. The study also tested the model's ability to predict bindings sites for transcription factors from maize.

68. * Zhao H, Tu Z, Liu Y, Zong Z, Li J, Liu H, Xiong F, Zhan J, Hu X, Xie W: **PlantDeepSEA, a deep learning-based web service to predict the regulatory effects of genomic variants in plants**. *Nucleic Acids Res* 2021, **49**:W523–W529.
Easy-to-use web server for the prediction of chromatin accessibility and how it is impacted by variation of the underlying DNA sequence.

69. ** Wrightsman T, Marand AP, Crisp PA, Springer NM, Buckler ES: **Modeling chromatin state from sequence across angiosperms using recurrent convolutional neural networks**. *Plant Genome* 2022, **15**, e20249.
Development of a deep learning approach to predict accessible chromatin regions and hypomethylated regions across several angiosperm species. The accuracy of these 'cross-species' models was comparable to or better than the accuracy of models trained and validated using a single species.

70. Zrimec J, Börlin CS, Buric F, Muhammad AS, Chen R, Siewers V, Verendel V, Nielsen J, Töpel M, Zelezniak A: **Deep learning suggests that gene expression is encoded in all parts of a co-evolving interacting gene regulatory structure**. *Nat Commun* 2020, **11**:6141.

71. Avsec Ž, Agarwal V, Visentin D, Ledsam JR, Grabska-Barwinska A, Taylor KR, Assael Y, Jumper J, Kohli P, Kelley DR: **Effective gene expression prediction from sequence by integrating long-range interactions**. *Nat Methods* 2021, **18**: 1196–1203.

72. Ranaweera T, Brown BNI, Wang P, Shiu S-H: **Temporal regulation of cold transcriptional response in switchgrass**. *bioRxiv* 2022. https://doi.org/10.1101/2022.08.01.502344.