Cell Press
REVIEWS

Opinion

Challenges of Translating Gene Regulatory Information into Agronomic Improvements

Nathan Springer,[1,*] Natalia de León,[2] and Erich Grotewold[3,*]

Improvement of agricultural species has exploited the genetic variation responsible for complex quantitative traits. Much of the functional variation is regulatory, in *cis*-regulatory elements and *trans*-acting factors that ultimately contribute to gene expression differences. However, the identification of gene regulatory network components that, when modulated, will increase plant productivity or resilience, is challenging, yet essential to provide increased predictive power for genome engineering approaches that are likely to benefit useful traits. Here, we discuss the opportunities and limitations of using data obtained from gene coexpression, transcription factor binding, and genome-wide association mapping analyses to predict regulatory interactions that impact crop improvement. It is apparent that a combination of information from these data types is necessary for the reliable identification and utilization of important regulatory interactions that underlie complex agronomic traits.

Highlights

The availability of an increasing number of plant genomes and the development of new high-throughput approaches is resulting in the identification of large collections of transcription factor-target gene interactions.

Natural variation is being utilized in a number of important crops to conduct genome-wide association studies to link DNA-sequence variation or changes in gene expression with phenotypes, often quantitative.

Many trait-associated DNA variants are located in gene regulatory regions, suggesting significant potential in linking gene regulatory information with breeding for complex phenotypes.

Fine-tuning the regulation of transcription factors could generate novel variation that would influence plant traits.

Gene Regulation and Genetic Variation

Control of gene expression is central to all aspects of cellular activity. Fundamental to this process are **transcription factors (TFs)** (see Glossary), proteins that bind to specific DNA sequences in the regulatory regions of the target genes that they control. TFs activate or repress transcription of specific sets of targets genes by recruiting, interacting, or modifying components of the transcriptional machinery [1]. TFs often function in a combinatorial fashion, allowing a discrete number of TFs to control the expression of a much larger number of target genes with unique temporal and spatial patterns [2]. Indeed, it is the concerted action of tens to hundreds of TFs tethered to regulatory regions through specific protein–protein and protein–DNA interactions (PDIs) that allow genes to be expressed with the appropriate expression patterns [2,3].

TFs can be hierarchically organized, such that one TF often controls the expression of a gene encoding another TF. Alternatively, some TFs regulate the expression of genes encoding structural proteins or metabolic enzymes, key players in many valuable **agronomic traits**. The hierarchical arrangement of TFs allows signals to be amplified, providing the information necessary for given sets of genes to be deployed with particular spatial and temporal patterns. Gene function is intimately linked to when and where genes are expressed. This information is hardwired in gene regulatory regions composed of proximal and distal *cis*-regulatory modules, formed by ***cis*-regulatory elements (CREs)** recognized by one or combinations of TFs (Figure 1). Proximal regulatory modules include the core promoter, a region of about 100 bp that flanks the transcription start site (TSS), and which is responsible for the assembly of the pre-initiation transcription complex and thus, the accurate initiation of transcription. In humans and other vertebrates, TF-binding sites are enriched within 1–2 Kb of the TSS, a situation that also appears to be the case in plants [4]. However, there are also many *cis*-regulatory modules located further away from the TSS that positively control gene expression. These distal CREs are often referred to as enhancers, if they are able to drive transcription independent of orientation, location, or relative distance with respect to the TSS (Figure 1). Enhancers are particularly dense in TF-binding sites, and the cooperative binding of multiple TFs is believed to be necessary to facilitate the nucleosome eviction necessary for enhancer activation. Thousands of enhancers are known in animals and are generally characterized by high chromatin accessibility, histone acetylation, the transcription of enhancer RNAs (eRNAs), and low DNA methylation [5]. However, significantly fewer enhancers have been identified in plants and, while it is clear that they share several properties with animals, the unifying characteristics of plant enhancers are still being determined using both chromatin profiling and comparative genomics approaches to identify conserved noncoding sequences [6–8].

[1]Department of Plant and Microbial Biology, University of Minnesota, St Paul, MN 55108, USA

[2]Department of Agronomy, University of Wisconsin, Madison, WI 56706, USA

[3]Department of Biochemistry and Molecular Biology, Michigan State University, East Lansing, MI 48824, USA

*Correspondence:
springer@umn.edu, grotewol@msu.edu

© 2019 Elsevier Ltd. All rights reserved.

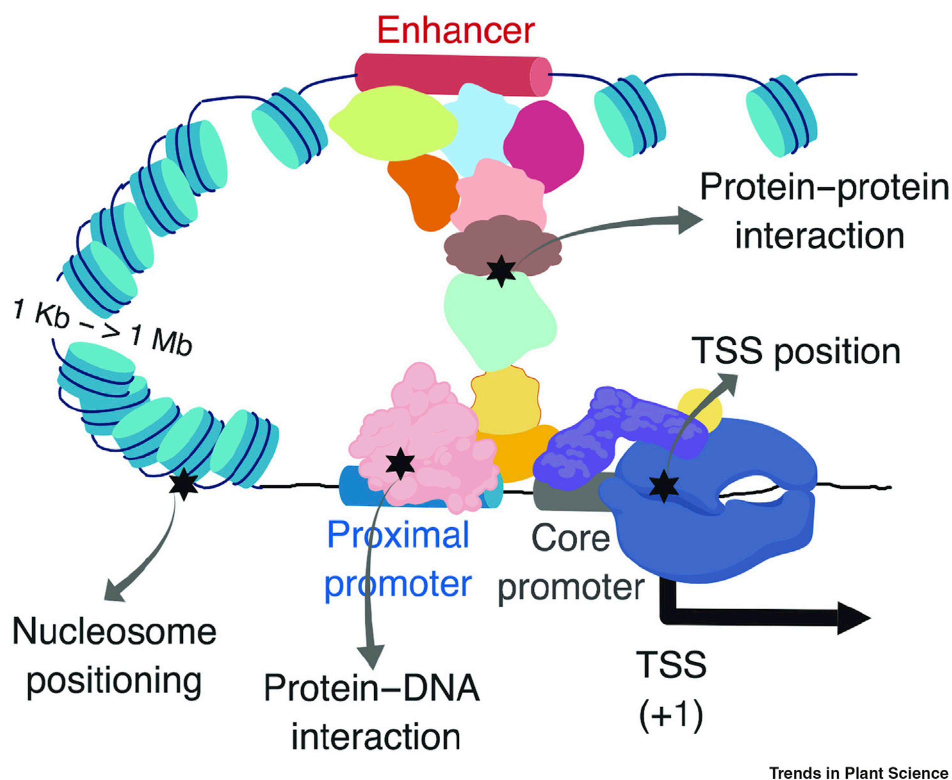

Figure 1. Main Components Required for Gene Expression that Can Be Affected by Natural Variation. Asterisks indicate the specific interactions required for gene regulation that can be affected by natural variation, including the position of the transcription start site (TSS); protein–protein interactions between transcription factors (TFs), TFs and cofactors, or TF/cofactors and components of the basal transcriptional machinery; interactions between TFs and DNA; and DNA-sequence changes that can influence nucleosome positioning.

Glossary

Agronomic trait: a complex phenotype that has agronomic importance because it influences yield or crop performance.

***Cis*-regulatory element (CRE)**: a small DNA sequence that serves as the binding site for a transcription factor.

***Cis*-regulatory variation**: differences in *cis*-regulatory elements between different accessions or landraces that may or may not influence transcription factor binding.

Gene coexpression network (GCN): a network in which the edges of the graph represent the correlation of expression between any two genes.

Gene regulatory network (GRN): a temporal and/or spatial manifestation of the gene regulatory grid that explains the interaction between all transcription factors and their target genes.

Genome-wide association studies (GWAS): study of the associations between phenotypes (traits) and specific regions in genomes.

Transcription factor (TF): a protein that binds DNA in a sequence-specific fashion and that contributes to the control of gene expression.

***Trans*-regulatory variation**: differences in transcription factors between different accessions or landraces that may influence the expression of their target genes.

Genetic variation describes the genotypic differences between individuals in a population and among populations, providing the diversity substrate required for selection. Genetic variation at a TF locus can influence TF abundance and/or TF activity (changes in TF protein sequence), potentially resulting in ***trans*-regulatory variation** for target genes. Alternatively, changes in the DNA sequence at the CREs of the target gene can result in ***cis*-regulatory variation**. A majority of the genetic variants that are associated with quantitative trait variation in maize are found within putative regulatory regions [9,10]. Other examples of genetic variation can affect protein–protein interactions that result in the formation of active regulatory complexes and chromosome looping or the organization of the nucleosomes that influence gene expression through chromatin structure (Figure 1).

Breeding for desirable traits often involves the selection of variant alleles with altered expression levels [11–15], either as a consequence of allelic variations in CREs or differences in the activity or expression of the corresponding TFs [16]. The relative influence of *cis*- and *trans*-regulatory variation can be assessed through expression quantitative trait locus (eQTL) mapping of regulatory diversity or through the analysis of expression levels progenies relative to parents [17]. Studies of regulatory variation in plant species have identified abundant *cis*- and *trans*-regulatory variation [18]. Lessons learnt in yeast and humans suggest that most expression variation arises from *trans*-acting eQTLs [19,20]; whether this is generally the case in plants as well remains to be determined. There have been differences in the frequency and magnitude of effects that have been discovered that could reflect biological variation among populations or technical differences in how regulatory variants were uncovered. An improved understanding of the specific DNA-sequence changes that underlie

CellPress
REVIEWS

trans- or *cis*-regulatory variation will be important for application of this knowledge to crop improvement through the creation of new alleles, or selection for specific combinations of alleles.

As we seek to harness regulatory variation to influence plant traits, it is desirable to focus on the modulation of the *trans*-acting factors. The observation that one or a small set of TFs can regulate multiple steps of a (metabolic) pathway [16,21] suggests that altering the activity or expression pattern of TFs could often provide greater influence over metabolism than changing CREs at a single gene of a pathway. In order to develop a roadmap for which TFs could influence specific pathways or traits, it is critical to develop **gene regulatory networks (GRNs)** that provide functional information about TF-target gene interactions. This type of knowledge will provide insights that can be used to improve the outcome of approaches that use targeted genomic modification to develop novel alleles with predictable impact on plant traits.

Avenues towards the Elucidation of GRNs

Transcriptional control, as opposed to RNA turnover or other aspects of gene expression, accounts for most of the reported variation in mRNA accumulation levels in animals [22]. Thus, it is reasonable to assume that most of the functional genetic variation in plants also has the primary consequence on gene transcription compared with RNA abundance, as seems to be the case in maize [23]. Central to understanding gene transcription are GRNs. GRNs are temporal and spatial manifestations of the gene regulatory grid that describes the complex web of functional PDIs necessary for the expression of all the genes in an organism [24]. Thus, a key step towards the characterization of GRNs is to determine the PDIs involving TFs and corresponding target genes. Approaches to determine PDIs can be divided into gene- and TF-centered [25]. Gene-centered approaches, such as yeast one-hybrid (Y1H), permit the identification of TFs that bind the promoter of a particular gene. The use of normalized TF Y1H libraries, rather than whole cDNA libraries, significantly enhances the proportion of 'true interactions', but eliminates the possibility of identifying novel DNA-binding proteins [26,27]. Y1H approaches have been successfully used to identify PDIs associated with *Arabidopsis* root development [28], cell wall synthesis [29], and cellular reprogramming [30], and maize phenolic compound regulation [31]. Yet, the identification of PDIs through Y1H has numerous limitations, including that only interactions involving the small (usually ~1 Kb) DNA bait region will be identified (Figure 2) and PDIs that require more than one TF or plant-specific TF post-translational modifications will be missed.

TF-centered approaches include chromatin-immunoprecipitation (ChIP) and DNA-affinity purification (DAP) assays, which, when coupled with massively parallel DNA sequencing (as in ChIP-Seq and DAP-Seq), allow the simultaneous identification of thousands of potential TF target genes [32,33]. The vast majority of the plant putative PDIs so far determined have been identified using TF-centered approaches and, for *Arabidopsis* alone, correspond to over 5 million interactions (see https://agris-knowledgebase.org/). Based on recent studies in other organisms [34], this number of PDIs likely corresponds to a small fraction of the interactions necessary to explain the *Arabidopsis* gene regulatory grid, emphasizing a significant challenge for crop plants with larger genomes and fewer molecular tools than *Arabidopsis*. ChIP-Seq permits the capture of *in vivo* TF-target interactions, but it can also identify target sequences indirectly bound by the TF (e.g., through another TF). Because it requires the availability of antibodies and relies on rather variable protein–DNA cross-linking and immunoprecipitation phenomena, ChIP-Seq is time-consuming and expensive to implement. In contrast, DAP-Seq allows a much higher throughput and is inexpensive, yet misses PDIs requiring multiple TFs or requiring TF post-translational modifications, and CREs identified by DAP-Seq can be within inaccessible chromatin regions *in vivo* and hence not be truly available for TF binding (Figure 2). The assessment of the biological relevance of ChIP-Seq or DAP-Seq enrichment at a particular locus often utilizes transcriptome profiling for loss-of-function alleles for the TF. However, in many crop plants, there are limited resources for genome-wide loss-of-function genotypes that could be used to confirm GRNs inferred from these TF-target interaction approaches.

A challenge in interpreting results obtained from gene- and TF-centered approaches is that identified PDIs frequently have little or no effect on mRNA accumulation when the expression of the TF involved

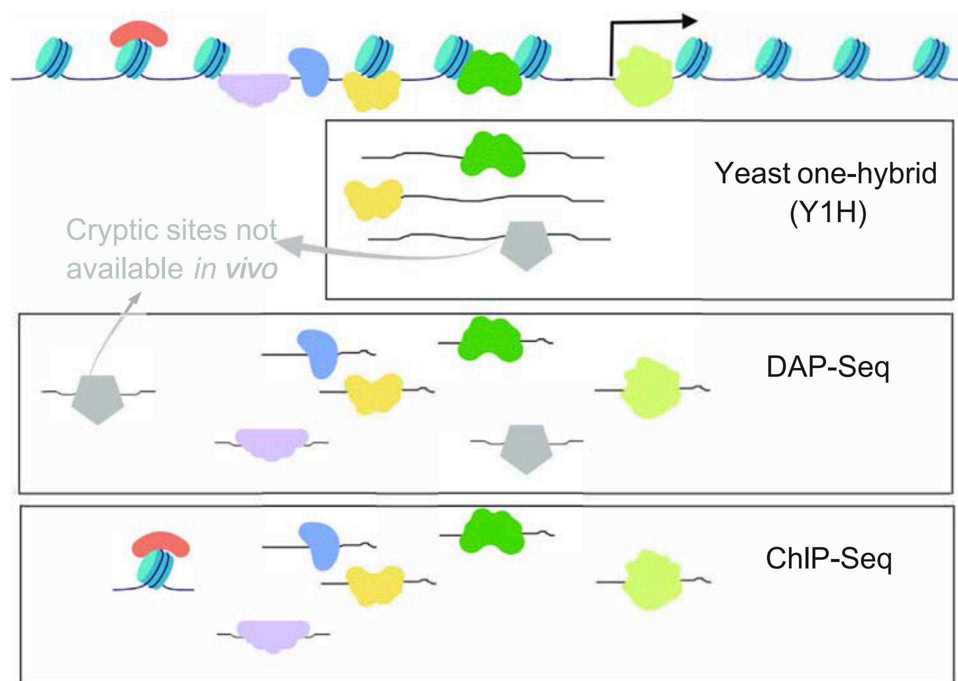

Figure 2. Protein–DNA Interactions Revealed by Gene- or Transcription Factor-Centered Approaches. Diagrammatic representation of the *cis*-regulatory region of a gene with nucleosomes (teal structures) and various DNA-binding transcription factors (colored structures). The arrow represents the transcription start site. The three squares indicate which protein–DNA interactions would be identified if conducting yeast one-hybrid (Y1H) studies with the represented portion of the gene regulatory region (usually ~1 Kb), by DNA-affinity purification (DAP)-Seq or by chromatin-immunoprecipitation (ChIP)-Seq. Note that *in vitro* methods such as Y1H and DAP-Seq can identify interactions with cryptic DNA sites that would be masked *in vivo*, for example, by nucleosomes.

is perturbed (e.g., [35–39]). While in some instances technical artifacts are responsible for the low overlap between target and differentially expressed genes, more often TF redundancy, poised and transient TF-target interactions [40], and regulation of the target gene by the TF in only a fraction of the cells sampled play a part [41]. For these reasons, recruitment of a TF to a particular gene regulatory region without a consequence on gene expression is interpreted as inconclusive from the perspective of the biological significance of that particular PDI [42,43]. It is also difficult to determine which PDIs are the ones that have the most significant impact, particularly since it is often the combinatorial interaction of multiple TFs with a regulatory region that determines the consequences on gene expression [44].

A complementary approach to inferring regulatory networks is through the use of statistical inference algorithms or machine learning techniques applied to gene expression data. In the simplest case, it is generally accepted that genes with very similar mRNA accumulation patterns (coexpression) are functionally related and are regulated by a similar mechanism involving shared TFs [45,46]. **Gene coexpression networks (GCNs)** and GRNs capturing specific spatio-temporal patterns of gene expression can be used as effective tools to identify genes involved in a specific pathway or process. However, the specific approaches used to construct GCNs (e.g., correlation method and of datasets used) can have significant impact on the resulting GCN [47]. Yet, GCNs have been useful in prioritizing genes underlying traits identified through **genome-wide association studies (GWAS)** in maize [48]. Also, computational inference methods have been employed to construct GRNs and find important genes and regulatory relationships involved in plant growth and developmental processes such as cell wall synthesis [29], regeneration [30], and root hair growth [49]. Several approaches have been

CellPress
REVIEWS

developed to utilize knowledge of annotated TFs to create putative GRNs from GCN data [50,51]. These coexpression-derived GRNs often exhibit some enrichment for known targets, but also contain many putative edges that are not observed in PDI datasets [52,53].

In summary, there are many approaches that can be utilized to uncover GRNs and assess their impact. Each approach has the potential to provide additional value and novel insights to such understanding, but false-positive and false-negative interactions can predispose subsequent steps and therefore complicate the generation of a robust GRN. There are challenges in finding the best strategy to integrate the output from the different methods used in elucidating complete GRNs. A key tool in the dissection and validation of GRNs has been the use of loss-of-function mutants in TFs. Improvements in our ability to scale-up approaches for the systematic genetic elucidation and assessment of individual TFs in crop species would greatly increase our understanding of the biologically relevant GRNs.

Opportunities and Challenges in Linking Natural Variation to GRNs

In some model species, GRNs can be tested and explored through the use of mutant libraries that can systematically probe the function of all genes [54]. Forward-genetic screens have identified mutants in many TFs in crop plants and, in some cases, these have been utilized to provide insights into critical GRNs in crop traits [38,39,55]. However, systematic genome-wide screens of TFs are not currently feasible in crop plants and forward-genetic approaches can be complicated by genetic redundancy and loss of viability of null alleles. This limits our ability to test the GRNs developed from TF-centered, gene-centered, or coexpression data. The study of natural variation present within crop species provides the opportunity to test some of the gene regulatory predictions offered by GRNs. In crop species, there is wide-spread variation of gene expression levels, and even gene content, among different varieties. This furnishes the opportunity to utilize natural variation that results in differences in TF presence/expression to assess whether putative target genes have altered expression. Also, in contrast with extreme loss- or gain-of-function mutations, the presence of different conformations of variant states, in a natural variation context, allow researchers to simultaneously test multiple combinations of TFs and target genes, especially in the case of diverse association panels with large population sizes. These approaches can also be extended to study differences in networks in wild and domesticated populations [56] or in closely related species [57].

Approaches focused on the study of natural variation have the benefit of simultaneously uncovering supported edges in GRNs and providing opportunities to utilize these alleles for crop improvement. The quantitative trait loci identified by GWAS often account for relatively small proportions of the variance for a trait, but once identified, variants can be enhanced or repressed via transgenic or editing approaches. This can result in the ability to generate allelic series with greater phenotypic effects than observed for alleles in natural populations. While gene expression levels could be influenced by changing CREs, TF variants have the potential for greater influence due to their control of multiple target genes [16,21]. Thus, focusing on characterizing the TFs with supported roles in GRNs could help uncover key control points influencing plant traits.

Pioneering studies on variations in plant and flower pigmentation resulted in the identification of the genes that encode pathway enzymes as well as key TFs, thanks to the ease in following phenotypes [58]. Our knowledge on the flavonoid pathway and its regulation across many plants provides evidence that GRNs can be elucidated through the analysis of natural variants and mutants, when phenotypes are easily scored. Many other studies have effectively utilized natural variation to probe the genetic basis for many quantitative traits through the analysis of bi-parental populations or association panels. A number of those studies have highlighted the importance of gene regulation as a mechanism to control the observable phenotypic variation, and examples include maize *teosinte branched1* (*tb1*) [11], *opaque2* (*o2*) [12], *ZmCCT* [13], *vegetative to generative transition 1* (*vgt1*) [14], and *teosinte glume architecture1* (*tga1*) [15]. Importantly, in many of these cases we have found natural variants that result in altered patterns of expression of TFs rather than true loss-of-function alleles, highlighting the potential for agronomic traits to arise through the manipulation of TFs.

CellPress
REVIEWS

Recent approaches to combine GWAS and GCNs suggest routes to integrate multiple evidence pieces to prioritize high quality candidate genes and to validate GRNs [48]. However, these studies will face a key limitation, since analyses of natural variation are restricted to the variants that are present in populations. Many crop species have populations of improved lines that have been subjected to intense selection pressure. Variation in key TFs may result in strong pleiotropic changes in phenotype that would be selected for, or against, and this can be exacerbated or ameliorated by the effect of the environment. This means that many surveys of natural variation will be limited to the variants that are tolerated in improved varieties and may have lost key variation that would be useful in exploring GRNs. This will also limit the potential to utilize GRN knowledge to create improved varieties through traditional plant breeding approaches. Advances in transformation efficiency and expanding the range of usable germplasm [59] open up the possibility to explore the effect of background variants in the presence of specific (especially if dominant in nature) perturbations across a wider range of natural diversity.

Applying GRNs for Crop Improvement

While there may be limits on the natural variation that is available in crops, there are other potential avenues to exploit GRN knowledge to develop improved varieties. Transgenic approaches or genome editing could be used for significant TFs to create novel alleles that are not represented in natural populations and to drive higher levels or novel patterns of expression. These changes may result in changes to the expression of targets that could modulate the outcome of pathways [60]. Alternatively, genome editing could be used to target distal regulatory regions of key TFs to create novel patterns or levels of expression of the endogenous gene. In both cases, they may provide alternatives to loss-of-function alleles typically generated by forward-genetics approaches and could result in novel traits. Recent studies on key tomato developmental regulators have shown that modulation of CREs can generate variants not seen in nature that can result in novel phenotypes [61]. Improvements in our ability to utilize chromatin profiles to identify distal regulatory elements [7,62,63] will guide efforts to create novel alleles of TFs that can be screened for improved agronomic performance. The potential to modulate regulation of TFs could mimic the creation of alleles observed in plant domestication of improvement [11–15].

It is worth noting that the application of GRNs to crop improvement will likely require different approaches for different types of traits. Some key metabolic traits may provide ideal starting points. There are many examples in which a single TF can regulate multiple structural genes of the same pathway and modulation of that TF can result in metabolic shifts in the pathway [21]. Given our understanding of enzymes and annotation of metabolic pathways it will likely be easier to characterize GRNs that are associated with activities of enzymes in metabolic pathways and modulate portions of these GRNs. In contrast, complex phenotypes such as yield or morphological traits may pose a greater challenge. There are not clear pathways or enzymes that necessarily affect these processes. Identifying GRNs and TFs that play critical roles in these traits, which are often very sensitive to environmental or genotype by environment interactions, will undoubtedly be more challenging. Mapping intermediate states, including component phenotypes, RNA transcripts, and/or metabolites, can improve resolution compared with mapping the resulting (and more complex) phenotype. The creation of novel variants for TFs that are associated with these processes could begin the process of screening for alleles or variants that provide improved performance or resilience.

Concluding Remarks and Future Perspectives

Despite substantial technological advances and continually improved understanding of cellular function, our ability to predict traits from genomes, in the context of variable environments, is limited due to the dynamic and complex networks that connect genotypes to phenotypes. Plant breeding operations continue to invest substantial resources in the conversion portion of their pipelines that introgresses QTL or transgenes into multiple genetic backgrounds. Background effects are known to have substantial impacts on the resulting phenotype and can limit the ability to deploy the QTL or transgene for improvement. The Quality Protein Maize project is a good example of the effort required to manage the potentially detrimental pleiotropic effect of a single mutant (*opaque2*) that otherwise

provides enhanced grain protein composition [64]. Traditionally, a forward breeding approach has also been preferred for the integration of transgenes to account for the potential dependency on endogenous alleles that can interact with the specific event in the context of the environment. For each of these situations, rational combinations of well-characterized functional genetic variants can enhance the efficiency of breeding outcomes [65]. A focus on strategies to better characterize network architecture and dynamics, including the role of functional versus genetic dependency and the preponderance of common versus rare regulatory variants, as well as a better understanding of what fraction of a genome plays a regulatory role (e.g., functioning as CREs) is expected to provide insights into efficient strategies to improve agronomic traits (see Outstanding Questions).

Outstanding Questions

Which method(s) capture best the protein–DNA interactions that are relevant for the regulation of agronomic traits?

To what extent is natural variation influencing gene regulation through mechanisms other than protein–DNA interactions, such as, for example, by changing the epigenome?

How informative are coexpression networks, from the perspective of targets of transcription factors and genes important for agronomic traits?

What is the functional *cis*-regulatory genome? In other words, what are all the DNA-sequences important for accurate gene expression?

Acknowledgments

We thank María Katherine Mejía-Guerra and Peng Zhou for valuable comments and ideas. This work was supported by funding from the National Science Foundation through grants IOS-1733633 and MCB-1822343.

References

1. Spitz, F. and Furlong, E.E.M. (2012) Transcription factors: from enhancer binding to developmental control. *Nat. Rev. Genet.* 13, 613
2. Brkljacic, J. and Grotewold, E. (2017) Combinatorial control of plant gene expression. *Biochim. Biophys. Acta Gene Regul. Mech.* 1860, 31–40
3. Wray, G.A. *et al.* (2003) The evolution of transcriptional regulation in eukaryotes. *Mol. Biol. Evol.* 20, 1377–1419
4. Yu, C.-P. *et al.* (2016) Positional distribution of transcription factor binding sites in *Arabidopsis thaliana. Sci. Rep.* 6, 25164
5. Long, H.K. *et al.* (2016) Ever-changing landscapes: transcriptional enhancers in development and evolution. *Cell* 167, 1170–1187
6. Weber, B. *et al.* (2016) Plant enhancers: a call for discovery. *Trends Plant Sci.* 21, 974–987
7. Oka, R. *et al.* (2017) Genome-wide mapping of transcriptional enhancer candidates using DNA and chromatin features in maize. *Genome Biol.* 18, 137
8. Freeling, M. and Subramaniam, S. (2009) Conserved noncoding sequences (CNSs) in higher plants. *Curr. Opin. Plant Biol.* 12, 126–132
9. Rodgers-Melnick, E. *et al.* (2016) Open chromatin reveals the functional maize genome. *Proc. Natl. Acad. Sci. U. S. A.* 113, E3177–E3184
10. Li, X. *et al.* (2012) Genic and nongenic contributions to natural variation of quantitative traits in maize. *Genome Res.* 22, 2436–2444
11. Studer, A. *et al.* (2011) Identification of a functional transposon insertion in the maize domestication gene *tb1. Nat. Genet.* 43, 1160–1163
12. Zhan, J. *et al.* (2018) *Opaque-2* regulates a complex gene network associated with cell differentiation and storage functions of maize endosperm. *Plant Cell* 30, 2425–2446
13. Yang, Q. *et al.* (2013) CACTA-like transposable element in *ZmCCT* attenuated photoperiod sensitivity and accelerated the postdomestication spread of maize. *Proc. Natl. Acad. Sci. U. S. A.* 110, 16969–16974
14. Salvi, S. *et al.* (2007) Conserved noncoding genomic sequences associated with a flowering-time quantitative trait locus in maize. *Proc. Natl. Acad. Sci. U. S. A.* 104, 11376–11381
15. Studer, A.J. *et al.* (2017) Selection during maize domestication targeted a gene network controlling plant and inflorescence architecture. *Genetics* 207, 755–765
16. Gray, J. and Grotewold, E. (2011) Transcription factors, gene regulatory networks and agronomic traits. In *Sustainable Agriculture and New Biotechnologies* (Benkeblia, N., ed,), pp. 65–94, CRC Press
17. Kliebenstein, D. (2009) Quantitative genomics: analyzing intraspecific variation using global gene expression polymorphisms or eQTLs. *Ann. Rev. Plant Biol.* 60, 93–114
18. Cubillos, F.A. *et al.* (2012) Lessons from eQTL mapping studies: non-coding regions and their role behind natural phenotypic variation in plants. *Curr. Opin. Plant Biol.* 15, 192–198
19. Grundberg, E. *et al.* (2012) Mapping cis- and trans-regulatory effects across multiple tissues in twins. *Nat. Genet.* 44, 1084
20. Albert, F.W. *et al.* (2018) Genetics of *trans*-regulatory variation in gene expression. *eLife* 7, e35471
21. Grotewold, E. (2008) Transcription factors for predictive plant metabolic engineering: are we there yet? *Curr. Opin. Biotechnol.* 19, 138–144
22. Pai, A.A. *et al.* (2015) The genetic and mechanistic basis for variation in gene regulation. *PLoS Genet.* 11, e1004857
23. Wallace, J.G. *et al.* (2014) Association mapping across numerous traits reveals patterns of functional variation in maize. *PLoS Genet.* 10, e1004845
24. Mejia-Guerra, M.K. *et al.* (2012) From plant gene regulatory grids to network dynamics. *Biochim. Biophys. Acta Gene Regul. Mech.* 1819, 454–465
25. Yang, F. *et al.* (2016) Establishing the architecture of plant gene regulatory networks. *Methods Enzymol.* 576, 251–304
26. Deplancke, B. *et al.* (2006) A gene-centered *C. elegans* protein-DNA interaction network. *Cell* 125, 1193–1205
27. Mitsuda, N. *et al.* (2010) Efficient yeast one-/two-hybrid screening using a library composed only of transcription factors in *Arabidopsis thaliana*. *Plant Cell Physiol.* 51, 2145–2151
28. Brady, S.M. *et al.* (2011) A stele-enriched gene regulatory network in the *Arabidopsis* root. *Mol. Syst. Biol.* 7, 459
29. Taylor-Teeples, M. *et al.* (2015) An *Arabidopsis* gene regulatory network for secondary cell wall synthesis. *Nature* 517, 571–575
30. Ikeuchi, M. *et al.* (2018) A gene regulatory network for cellular reprogramming in plant regeneration. *Plant Cell Physiol.* 59, 770–782
31. Yang, F. *et al.* (2017) A maize gene regulatory network for phenolic metabolism. *Mol. Plant* 10, 498–515
32. Furey, T.S. (2012) Chip–seq and beyond: new and improved methodologies to detect and characterize

protein–DNA interactions. *Nat. Rev. Genet.* 13, 840–852
33. O'Malley, R.C. *et al.* (2016) Cistrome and epicistrome features shape the regulatory DNA landscape. *Cell* 165, 1280–1292
34. Ouma, W.Z. *et al.* (2018) Topological and statistical analyses of gene regulatory networks reveal unifying yet quantitatively different emergent properties. *PLoS Comput. Biol.* 14, e1006098
35. Zeller, K.I. *et al.* (2006) Global mapping of c-myc binding sites and target gene networks in human B cells. *Proc. Natl. Acad. Sci. U. S. A.* 103, 17834
36. Liu, S. *et al.* (2015) Negative regulation of ABA signaling by WRKY33 is critical for *Arabidopsis* immunity towards *Botrytis cinerea* 2100. *eLife* 4, e07295
37. Morohashi, K. and Grotewold, E. (2009) A systems approach reveals regulatory circuitry for *Arabidopsis* trichome initiation by the *gl3* and *gl1* selectors. *PLoS Genet.* 5, e1000396
38. Morohashi, K. *et al.* (2012) A genome-wide regulatory framework identifies maize *pericarp color1* controlled genes. *Plant Cell* 24, 2745–2764
39. Eveland, A.L. *et al.* (2014) Regulatory modules controlling maize inflorescence architecture. *Genome Res.* 24, 431–443
40. Para, A. *et al.* (2014) Hit-and-run transcriptional control by bZIP1 mediates rapid nutrient signaling in *Arabidopsis*. *Proc. Natl. Acad. Sci. U. S. A.* 111, 10371–10376
41. Swift, J. and Coruzzi, G.M. (2017) A matter of time — how transient transcription factor interactions create dynamic gene regulatory networks. *Biochim. Biophys. Acta Gene Regul. Mech.* 1860, 75–83
42. Jiang, S. and Mortazavi, A. (2018) Integrating chip-seq with other functional genomics data. *Brief. Funct. Genomics* 17, 104–115
43. Banks, C.J. *et al.* (2016) Functional transcription factor target discovery via compendia of binding and expression profiles. *Sci. Rep.* 6, 20649
44. Grossman, S.R. *et al.* (2017) Systematic dissection of genomic features determining transcription factor binding and enhancer function. *Proc. Natl. Acad. Sci. U. S. A.* 114, E1291–E1300
45. Eisen, M.B. *et al.* (1998) Cluster analysis and display of genome-wide expression patterns. *Proc. Natl. Acad. Sci. U. S. A.* 95, 14863–14868
46. Haynes, B.C. *et al.* (2013) Mapping functional transcription factor networks from gene expression data. *Genome Res.* 23, 1319–1328
47. Huang, J. *et al.* (2017) Construction and optimization of a large gene coexpression network in maize using RNA-Seq data. *Plant Physiol.* 175, 568
48. Schaefer *et al.* (2018) Integrating coexpression networks with GWAS to prioritize causal genes in maize. *Plant Cell* 30, 2922
49. Shibata, M. *et al.* (2018) *GTL1* and *DF1* regulate root hair growth through transcriptional repression of *ROOT HAIR DEFECTIVE 6-LIKE 4* in *Arabidopsis*. *Development* 145, dev159707
50. Song, L. *et al.* (2012) Comparison of co-expression measures: mutual information, correlation, and model based indices. *BMC Bioinformatics* 13, 328
51. Schaefer, R.J. *et al.* (2017) Unraveling gene function in agricultural species using gene co-expression networks. *Biochim. Biophys. Acta Gene Regul. Mech.* 1860, 53–63
52. Walley, J.W. *et al.* (2016) Integration of omic networks in a developmental atlas of maize. *Science* 353, 814–818
53. Huang, J. *et al.* (2018) Distinct tissue-specific transcriptional regulation revealed by gene regulatory networks in maize. *BMC Plant Biol.* 18, 111
54. Scherens, B. and Goffeau, A. (2004) The uses of genome-wide yeast mutant collections. *Genome Biol.* 5, 229
55. Bolduc, N. *et al.* (2012) Unraveling the KNOTTED1 regulatory network in maize meristems. *Genes Dev.* 26, 1685–1690
56. Koenig, D. *et al.* (2013) Comparative transcriptomics reveals patterns of selection in domesticated and wild tomato. *Proc. Natl. Acad. Sci. U. S. A.* 110, E2655
57. Hu, G. *et al.* (2016) Evolutionary conservation and divergence of gene coexpression networks in *Gossypium* (cotton) seeds. *Genome Biol. Evol.* 8, 3765–3783
58. Grotewold, E. (2006) The genetics and biochemistry of floral pigments. *Annu. Rev. Plant Biol.* 57, 761–780
59. Lowe, K. *et al.* (2016) Morphogenic regulators *BABY BOOM* and *WUSCHEL* improve monocot transformation. *Plant Cell* 28, 1998–2015
60. Francis, D. *et al.* (2017) Challenges and opportunities for improving food quality and nutrition through plant biotechnology. *Curr. Opin. Biotech.* 44, 124–129
61. Rodriguez-Leal, D. *et al.* (2017) Engineering quantitative trait variation for crop improvement by genome editing. *Cell* 171, 470–480
62. Lu, Z. *et al.* (2018) Identification of *cis*-regulatory elements by chromatin structure. *Curr. Opin. Plant Biol.* 42, 90–94
63. Maher, K.A. *et al.* (2018) Profiling of accessible chromatin regions across multiple plant species and cell types reveals common gene regulatory principles and new control modules. *Plant Cell* 30, 15–36
64. Jia, M. *et al.* (2013) Identification and characterization of lysine-rich proteins and starch biosynthesis genes in the opaque2 mutant by transcriptional and proteomic analysis. *BMC Plant Biol.* 13, 60
65. Wallace, J.G. *et al.* (2018) On the road to breeding 4.0: unraveling the good, the bad, and the boring of crop quantitative genomics. *Ann. Rev. Genet.* 52, 421–444