TITLE: Improving plant gene regulatory network inference by integrative analysis of multi-omics and high resolution datasets

Yichun Qian[1], Shao-shan Carol Huang[1]

Address:

[1]Center for Genomics and Systems Biology, Department of Biology, New York University, 12 Waverly Pl, New York, NY 10003, U.S.A.

Corresponding author: Shao-shan Carol Huang (s.c.huang@nyu.edu)

ABSTRACT

Gene regulatory networks (GRNs) model the interactions between gene expression regulators and their targets that mediate a myriad of biological functions. Constructing GRNs that integrate multiple data types at increased resolution is improving our understanding of the complex regulatory mechanisms controlling different biological processes in plants. Going beyond transcription factor binding and transcriptome profiles, GRNs that incorporate multiple data types, including chromatin accessibility and long-range chromatin interaction, TF binding site motifs, microRNA, ribosome-associated RNA, and proteomic profiles, were constructed for several cell types and multiple species. The rise of single-cell RNA-seq applications in plants opens up possibilities for studying cell type-specific GRNs in the processes of cell differentiation, development and responses to the environment. Applications of high-throughput reporter assays and genome editing technologies allow large-scale validation of GRNs. Future advances in refining plant GRNs will most likely involve integration of multi-omics single-cell data and methods for cross-species model translation.

MAIN TEXT

Introduction

A gene regulatory network (GRN) depicts the interactions between gene expression regulators and their targets, as well as how these interactions affect the expression levels of RNA and proteins. GRNs provide a holistic view of the interactions between molecules that mediate biological processes and phenotypic traits (Figure 1). Many computational methodologies have been developed to build comprehensive GRNs (reviewed in [1,2]), and recent systems biology studies have made significant progress in delineating plant GRNs involved in various aspects of plant development [3*,4], environmental responses [5**,6], and coordination between growth and defense [7,8]. However, there are still significant challenges in constructing GRNs that recapitulate the complex and heterogeneous biological processes in the plant. In this review, we focus on recent progress in addressing these challenges in constructing plant GRNs by integrating multiple data types and increasing the spatial and temporal resolution, as well as experimental approaches that enable the critical task of GRN validation.

Integration of multiple data types in GRN

A majority of studies on plant GRNs focused on how changes in transcription factor (TF) binding events affect target gene expression, under the premise that changes in target gene levels could be predictive of changes in the relevant biological processes. However, it is difficult to predict changes in biological processes, cellular structures or phenotypes solely based on observed changes in the levels of mRNA transcripts or proteins for several reasons. First, mRNA levels are not necessarily correlated with protein abundance [9,10]. Secondly, transcriptional compensation between paralogous genes and functional redundancy between proteins are prevalent in plants, in which loss of certain transcripts or proteins can be compensated by proteins with overlapping functions [11,12]. For example, mutating two master regulators involved in tomato fruit ripening resulted in partial non-ripening phenotypes in tomato fruits, suggesting this important biological process may be regulated by redundant GRNs [13]. Furthermore, post-transcriptional regulation by microRNA (miRNA) and post-translational modifications also have significant impact on a broad range of biological processes [14-16]. Therefore, in order to obtain a holistic picture of the biological processes in plant cells, GRNs need to be constructed beyond transcriptome profiles.

The development of new experimental approaches for genome-wide high resolution identification of *cis*-regulatory elements (CREs), accessible chromatin regions (ACRs), and chromatin architecture features has made it possible to discover and integrate new features in plant GRNs (Table 1). These methods are particularly valuable for plants since the rapid sequence evolution of plant genomes make it challenging to use phylogenetic footprinting methods that were developed in mammalian genomes for regulatory element identification [17,18]. Integrative analysis of epigenome dynamics (H3K27ac profile and chromatin accessibility) and transcriptome dynamics in the developing flowers of *Arabidopsis thaliana* revealed that DNase I hypersensitive sites (DHS) in distal intergenic regions were predictive of

active enhancers and distinct sets of TF motifs were enriched in stage-specific enhancers [19*], allowing the construction of GRNs that are stage-specific rather than static. The distal DHS showed features of active enhancers because genes with both distal and proximal DHS had higher expression levels than genes with distal or proximal DHS only, and the accessibility dynamics of these regions among different floral developmental stages correlated with stage-specific gene expression [19*]. Changes in genome-wide chromatin accessibility between the stem cells of the shoot apical meristem (SAM) and differentiated leaf mesophyll cells in *Arabidopsis* were used to infer TF regulatory networks specific to these two cell types [20*]. Incorporation of chromatin interaction data could link long-range CREs to target genes, especially in large plant genomes. A genome-wide study of CREs in maize seedlings [21**] revealed that ACRs located far away from genes (distal ACRs; dACRs) were prevalent in the maize genome and that dACRs formed chromatin loops to their target genes and acted as enhancers.

In addition to transcriptome profiles, multiple -omics scale data types have been continuously integrated into mRNA-centric GRNs to improve network inference. In maize, GRN derived from integrating transcriptome, proteome and phosphoproteome data significantly outperformed GRNs derived from a single data type [9]. Translatome data generated by TRAP-seq (Table 1) can be used as an alternative to proteome data [22,23], complementing certain limitations of proteomics, i.e. low coverage due to lowly expressed or unstable proteins. Integrative analysis of multi-level epigenome, transcriptome and translatome data from *Arabidopsis* seedlings in response to hypoxia revealed comprehensive gene regulation dynamics coordinating chromatin accessibility, transcription and translation processes that occur in the nuclei and cytoplasm under environmental stress [23]. Analysis of floral gene regulatory network in *Arabidopsis* that included genome-wide TF binding, mRNA and miRNA expression data identified 568 feed-forward loop (FFL) motifs in which a master TF targeted both the miRNA and miRNA-targeted TFs [3*]. In a coherent FFL, the floral master regulator SEP3 binds to the upstream regions of *MIR319a*, *TCP4* and *TCP10*, while *MIR319a* in turn targets *TCP4* and *TCP10. MIR319a* is a critical component in this FFL that regulates petal growth by controlling the timing of *TCP4* and *TCP10* activation.

The diversity in the plant kingdom provides an advantage to study the evolution and conservation of GRN that mediates specific biological processes. By tracing the evolution of a single phenotypic trait among natural populations within the same species, as well as between closely and distantly related species, conservation characteristics of plant GRNs could be characterized. Analysis of dACRs in thirteen angiosperms showed that the majority of dACR sequences and their chromatin environment were conserved between species [24]. Transposable element proliferation contributed to the species-specific distribution of dACRs, and partially explained why species with larger genome sizes contained a higher proportion of dACRs [24]. Incorporating environmental conditions into evolutionary genomics analysis revealed conserved GRNs involved in environmental responses. Survey of RNA transcripts responsive to submergence in four dryland-adapted and flood-resilient angiosperm species identified 68 submergence-up-regulated families (SURFs), and motif analysis of promoters and accessible chromatin regions near the SURFs reported enrichment of four TF binding motifs (HRPE,

bHLH, WRKY, MYB). Putative regulatory networks for each species were built based on the presence of these motifs at each SURF gene in each species, and comparison of the networks across species suggested species-specific and conserved functionalities of these TF motifs in flood response [5**]. Interestingly, although flooding response circuitry was found in dryland-adapted species and wetland crop, it showed higher degree of activation in the wetland crop in response to submergence [5**]. Comparative analysis of natural selection strength on over 15,000 transcripts in two populations of rice under drought and wet conditions reported that earlier flowering and higher expression of early flowering regulator *OsMADS18* were strongly selected under drought conditions, and selection strength was weaker for genes with high connectivity in GRNs, higher number of CREs and transcriptional regulators [25]. The evolutionary characteristics of GRNs could provide insights for enhancing plant resilience to environmental changes.

Increasing the resolution of GRN

A plant is a complex system consisting of different organs, tissues and cell types at various developmental stages and with distinct sets of biological processes. Therefore, GRNs need to be built to reflect the different developmental stages, the diverse internal and external molecular environments, and the positional context of each cell type. The application of cell type-specific and single cell approaches and the addition of temporal and spatial information have increased the resolution of plant GRNs and improved our understanding of biological processes in plants.

Cell type-specific genetic markers allow isolation of cell types of interest and construction of cell type-specific GRNs. INTACT-ATAC-seq (Table 1) of *Arabidopsis* SAM stem cells and leaf mesophyll cells identified ACRs and enriched motifs specific to each cell type [20*]. By examining TF motif enrichment in differentially accessible ACRs of these two cell types, cell type-specific TFs and their target genes were predicted to create TF regulatory networks specific to each cell type [20*]. These GRNs were further expanded by adding new cell type-specific TF binding events discovered by an Ensemble motif-mapping approach [26]. INTACT-ATAC-seq was also used to analyze hair and non-hair cells in the *Arabidopsis* root, which reported a root hair cell specific MYB-driven regulatory module that controlled cell fate and response to abiotic stress, such as water and phosphate starvation [27]. By immunoprecipitation of epitope-tagged ribosomal subunits expressed under the control of domain specific promoters, domain-specific translatome profiles and gene co-expression networks were generated for nine different SAM and leaf domains in *Arabidopsis* [22]. Interestingly, leaf marginal domain was found to share high gene expression similarity with the unrelated rib meristem domain, supporting the parallel morphogenesis hypothesis between leaf and SAM [22].

One shortcoming of using markers is that cell type markers must be defined *a priori* [20,22,27,61], limiting the identification of new cell types or transient cell states and resulting in incomplete cell type-specific GRNs. Single-cell RNA-seq (scRNA-seq) provides a marker-free approach to characterize cell types and cell states and to construct GRNs. scRNA-seq profiling of *Arabidopsis* root cells discovered genes with cell type-specific expression that could be used

as new markers for cell type identification [28-31*], and generated developmental trajectories for multiple cell types such as the endodermal and epidermal cells [28,29]. Rather than grouping cells by collection time in the experiments, the trajectories placed individual cells along a continuous "pseudotime" path that represents the developmental and differentiation progression of the cells, allowing inference of more refined cell type- and stage-specific transcriptional regulators [30**] and GRNs [31*]. Examining the co-occurrence of TF expression and motif enrichment in specific cell clusters revealed putative TFs that drove cell cluster-specific gene expression, as well as TFs that drove gene expression at early and late stages along the developmental trajectories in cortex, endodermal and root hair cells [30**]. A GRN regulating trichoblast differentiation process was built for 239 TFs that were dynamically expressed across the pseudotime trajectory using the SCODE algorithm [31*,62]. SCODE used ordinary differential equations (ODE) to model the pseudotime expression dynamics of the TFs, and inferred TF regulatory networks by optimizing the parameters of the ODE to reconstruct the observed expression data [62]. This GRN predicted key TFs, such as ATHB-20, that might be involved in root development, and revealed a negative feedback relation between TFs at the end of trajectory and TFs in the meristem [31*]. scRNA-seq also makes it possible to characterize cell type-specific responses to environmental changes, as demonstrated by scRNA-seq analysis of *Arabidopsis* root following heat stress [30**]. Different biological processes were enriched for genes that showed cluster-specific expression changes in response to heat shock, such as ribosome-associated and RNA methylation in hair cells, cell wall organization and biogenesis in stele cells, and nitrogen and anion transport in endodermis cells [30**]. Such datasets provide a foundation for constructing GRNs that underlie cell type-specific responses to environmental stimuli.

Incorporating time and spatial information provides additional dimensions to investigate GRNs. In *Arabidopsis*, temporal regulation on nitrogen response genes was uncovered by applying dynamic factor graphs (DFG) [63] to time-series transcriptome datasets from *Arabidopsis* shoots in response to nitrogen signaling [32*]. In a state-space modeling framework, DFG learned a function that determined the target gene expression at each time point from the expression of a set of TFs at previous time points [32*]. This function represented the influence of TFs on the target genes, giving rise to a putative GRN that was pruned based on validation data from TARGET assays (Table 1). The pruned and validated GRN predicted the dynamic relationship between 155 TFs and 608 nitrogen-response genes [32*]. Using a barcoded array-based spatial transcriptome technique [33], meristem micro domain specific genes and biosynthetic pathways involved in different meristem developmental stages were identified and visually localized in the inflorescence meristem of *Arabidopsis* [34*]. However, the current resolution of this method (200 μm) is significantly larger than the size of plant cells (30 μm to 100 μm in *Arabidopsis* [35]), so information regarding rare cell types or domains consisting of single cell layers may be missed.

Experimental validation of GRN

Once a GRN is constructed, we need robust tools to validate the connections in the network. Two widely used validation approaches are reporter assays and targeted mutagenesis.

Multiple reporter assays have been applied to validate GRNs in plants. A cell-based transient TF perturbation system named TARGET (Table 1) was used to validate *in vivo*, direct TF-target interactions in nitrogen response in *Arabidopsis* protoplasts [36]. In maize, the regulatory functions of DHSs associated with conserved noncoding sequences were validated by GFP reporter expression in protoplasts [37]. One caveat of using protoplasts is that the protoplast preparation process may induce changes in the endogenous molecular environment of the cell. In addition to genome-wide enhancer mapping, STARR-seq (Table 1) can be used to validate potential enhancer candidates [21**], although the strength of enhancers is difficult to evaluate due to the lack of endogenous genome environment for the fragments being tested. A rapid validation method for enhancers was developed in tobacco using an agroinfiltration luciferase reporter assay, which allowed the relative strength of different enhancers to be quantitatively compared through bioluminescence signals [38].

Targeted mutagenesis can be used to test the regulatory effect of TFs and CREs on specific phenotypic traits. Recent studies in tomato used CRISPR/Cas9 genome editing to generate triple mutants of transcriptional regulators controlling flowering time (*SP5G*), growth termination (*SP*) and stem length (*SlER*), resulting in dwarf tomato plants without compromising yield [39,40]. CRISPR/Cas9 mutagenesis at the promoter region of *SlCLV3* created quantitative variation in tomato locule numbers [41*]. However, the phenotypic effect is not strictly correlated with the magnitude of disruption in the promoter region, further suggesting the complex relationship between GRNs and *in planta* biological processes [41*]. The development of multigene transcriptional activation systems in plants, such as the multiplexed CRISPR-Act2.0 [42], has the potential for large-scale, systematic validation of GRN.

Future perspectives

Currently, most of the plant GRNs were constructed based on whole tissue sections and bulk measurements. However, these approaches cannot capture the dynamics of cell interactions, such as the transport of small molecules and proteins across cells. These interactions play critical roles in coordinating gene expression programs and biological processes that give rise to phenotypic traits in plants. Therefore, plant GRN models need to be improved by integrating multi-omic, cell type-specific datasets, with the ultimate goal to understand and create precise phenotypic traits.

Single cell approaches have been developed and used in mammalian cells to collect information on chromatin accessibility, *cis*-regulation, transcriptome and proteome [43-46]. Cross-platform integration of multi-omic single cell datasets significantly improved the resolution of cell type identification [47,48]. It is also possible to simultaneously profile multiple modalities, such as chromatin architecture, epigenome, transcriptome, and proteome, in the same

cell [10,49-50]. Although these high resolution and integrative single cell approaches hold great promise to advance plant GRN research, they need to be adjusted to accommodate plant-specific features and challenges, such as the presence of cell wall, chloroplasts, vacuole and secondary metabolites.

Many GRNs have been studied in model plant species, so it is important to investigate the extent at which the GRN information could be generalized to non-model species. The challenge lies in the selection of translatable features and representative phenotypes, because regulatory mechanisms that underlie the same biological process may differ in different plant species. For example, the gene networks that regulate ovule initiation are distinct in *Arabidopsis* and tomato [51]. Comparing gene expression programs between species at single cell resolution alleviates the confounding variation in tissue and organ anatomy, potentially allowing more accurate cross-species comparisons [52]. Machine learning methods for translating animal models to human patients [53], which integrate phenotypes and multiple data types and incorporate cross-species differences, could inspire new approaches for translating GRNs in plants.

Table 1. Experimental techniques for identification and validation of GRN.

Method	Description	GRN feature identified	References
STARR-Seq (self-transcribing active regulatory region sequencing)	Transfect cells with a construct that contains a minimum promoter upstream of a sequence of interest, followed by sequencing of cellular RNA	Enhancer elements	[21**,54,55]
DAP-seq (DNA affinity purification sequencing)	Affinity purification of genomic DNA by *in vitro* expressed TF followed by sequencing	Transcription factor binding sites	[21**,56]
ATAC-seq (assay for transposase-accessible chromatin sequencing)	Digestion of chromatin at by Tn5 transposase, followed by sequencing	Accessible chromatin regions	[5**,21**,23,27,57,58]
HiChIP	Chromatin immunoprecipitation (ChIP) of factor-directed chromatin contacts, followed by sequencing	Protein mediated chromatin interaction	[21**,59]
TRAP-seq (translating ribosome affinity purification followed by RNA-seq)	Affinity purification of RNA bound by an epitope-tagged ribosomal protein, followed by RNA sequencing	Translating RNA	[5**,22,23]
INTACT (isolation of nuclei tagged in specific cell types)	Transgenic expression of nuclear envelope proteins followed by affinity purification of the tagged nuclei	Nuclei from specific tissue or cell type	[60]
INTACT-ATAC-seq	ATAC-seq with INTACT isolated nuclei	Cell type-specific accessible chromatin regions	[5**,20*,23,27]
TARGET (Transient assay reporting genome-wide effects of transcription factors)	Transient expression of TF in protoplasts followed by RNA sequencing	Target genes activated by a TF	[36]

Acknowledgments

This work was supported by an National Science Foundation grant to S.C.H (IOS-1916804).

References and recommended reading

* of special interest

** of outstanding interest

1. 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.
2. Delgado FM, Gomez-Vela F: **Computational methods for Gene Regulatory Networks reconstruction and analysis: A review**. *Artif Intell Med* 2019, **95**:133-145.
*3. Chen D, Yan W, Fu LY, Kaufmann K: **Architecture of gene regulatory networks controlling flower development in *Arabidopsis thaliana***. *Nat Commun* 2018, **9**:4534.

This study revealed that miRNA participated in the GRN controlling *Arabidopsis* flower development by forming a feed forward loop (FFL) with other regulators.

4. Doroshkov AV, Konstantinov DK, Afonnikov DA, Gunbin KV: **The evolution of gene regulatory networks controlling *Arabidopsis thaliana* L. trichome development**. *BMC Plant Biol* 2019, **19**:53.
5. Reynoso MA, Kajala K, Bajic M, West DA, Pauluzzi G, Yao AI, Hatch K, Zumstein K, Woodhouse M, Rodriguez-Medina J, et al.: **Evolutionary flexibility in flooding response circuitry in angiosperms. *Science* 2019, **365**:1291-1295.

This study illustrated how comparative genomics contribute to constructing GRNs. By comparing genome-wide chromatin accessibility and RNA transcriptome under transcriptional and posttranscriptional regulation from four angiosperm species, conserved GRNs in response to submergence were identified. Finding the molecular basis of these evolutionarily conserved traits could reveal mechanisms key to the survival of plants and is crucial in translating GRN models across species.

6. Van Moerkercke A, Duncan O, Zander M, Simura J, Broda M, Vanden Bossche R, Lewsey MG, Lama S, Singh KB, Ljung K, et al.: **A MYC2/MYC3/MYC4-dependent transcription factor network regulates water spray-responsive gene expression and jasmonate levels**. *Proc Natl Acad Sci U S A* 2019, **116**:23345-23356.
7. Li B, Tang M, Caseys C, Nelson A, Zhou M, Zhou X, Brady SM, Kliebenstein DJ: **Epistatic Transcription Factor Networks Differentially Modulate *Arabidopsis* Growth and Defense**. *Genetics* 2020, **214**:529-541.
8. Sanz-Carbonell A, Marques MC, Martinez G, Gomez G: **Dynamic architecture and regulatory implications of the miRNA network underlying the response to stress in melon**. *Rna Biology* 2020, **17**:292-308.
9. Walley JW, Sartor RC, Shen Z, Schmitz RJ, Wu KJ, Urich MA, Nery JR, Smith LG, Schnable JC, Ecker JR, et al.: **Integration of omic networks in a developmental atlas of maize**. *Science* 2016, **353**:814-818.
10. Darmanis S, Gallant CJ, Marinescu VD, Niklasson M, Segerman A, Flamourakis G, Fredriksson S, Assarsson E, Lundberg M, Nelander S, et al.: **Simultaneous Multiplexed Measurement of RNA and Proteins in Single Cells**. *Cell Rep* 2016, **14**:380-389.
11. Rodriguez-Leal D, Xu C, Kwon CT, Soyars C, Demesa-Arevalo E, Man J, Liu L, Lemmon ZH, Jones DS, Van Eck J, et al.: **Evolution of buffering in a genetic circuit controlling plant stem cell proliferation**. *Nat Genet* 2019, **51**:786-792.

12. Li R, Qiu Z, Wang X, Gong P, Xu Q, Yu QB, Guan Y: **Pooled CRISPR/Cas9 reveals redundant roles of plastidial phosphoglycerate kinases in carbon fixation and metabolism**. *Plant J* 2019, **98**:1078-1089.
13. Gao Y, Zhu N, Zhu X, Wu M, Jiang CZ, Grierson D, Luo Y, Shen W, Zhong S, Fu DQ, et al.: **Diversity and redundancy of the ripening regulatory networks revealed by the fruitENCODE and the new CRISPR/Cas9 CNR and NOR mutants**. *Hortic Res* 2019, **6**:39.
14. Wang K, Yang Z, Qing D, Ren F, Liu S, Zheng Q, Liu J, Zhang W, Dai C, Wu M, et al.: **Quantitative and functional posttranslational modification proteomics reveals that TREPH1 plays a role in plant touch-delayed bolting**. *Proc Natl Acad Sci U S A* 2018, **115**:E10265-E10274.
15. Szweykowska-Kulinska Z, Jarmolowski A: **Post-transcriptional Regulation of MicroRNA Accumulation and Function: New Insights from Plants**. *Mol Plant* 2018, **11**:1006-1007.
16. Mateos JL, de Leone MJ, Torchio J, Reichel M, Staiger D: **Beyond Transcription: Fine-Tuning of Circadian Timekeeping by Post-Transcriptional Regulation**. *Genes* 2018, **9**:616
17. Reineke AR, Bornberg-Bauer E, Gu J: **Evolutionary divergence and limits of conserved non-coding sequence detection in plant genomes**. *Nucleic Acids Res* 2011, **39**:6029-6043.
18. Hupalo D, Kern AD: **Conservation and functional element discovery in 20 angiosperm plant genomes**. *Mol Biol Evol* 2013, **30**:1729-1744.
*19. Yan W, Chen D, Schumacher J, Durantini D, Engelhorn J, Chen M, Carles CC, Kaufmann K: **Dynamic control of enhancer activity drives stage-specific gene expression during flower morphogenesis**. *Nat Commun* 2019, **10**:1705.

This study showed that stage-specific gene expression during *Arabidopsis* flowering development was associated with dynamic enhancers predicted by differential chromatin accessibility. Searching for TF motifs enriched in these enhancers identified potential TFs that might drive stage-specific gene expression.

*20. Sijacic P, Bajic M, McKinney EC, Meagher RB, Deal RB: **Changes in chromatin accessibility between Arabidopsis stem cells and mesophyll cells illuminate cell type-specific transcription factor networks**. *Plant J* 2018, **94**:215-231.

This study incorporated differential chromatin accessibility profiles to infer cell type-specific GRN in *Arabidopsis* SAM stem cells and leaf mesophyll cells.

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

This study characterized the prevalence of distal accessible chromatin regions in the maize genome, and predicted and validated their functions as enhancers at genome-wide scale. Considering the dynamic control of enhancers on gene expression, the existence of genome-wide, long-range *cis*-regulatory elements provides new information for constructing GRNs.

22. Tian C, Wang Y, Yu H, He J, Wang J, Shi B, Du Q, Provart NJ, Meyerowitz EM, Jiao Y: **A gene expression map of shoot domains reveals regulatory mechanisms**. *Nat Commun* 2019, **10**:141.
23. Lee TA, Bailey-Serres J: **Integrative Analysis from the Epigenome to Translatome Uncovers Patterns of Dominant Nuclear Regulation during Transient Stress**. *Plant Cell* 2019, **31**:2573-2595.
24. Lu Z, Marand AP, Ricci WA, Ethridge CL, Zhang X, Schmitz RJ: **The prevalence, evolution and chromatin signatures of plant regulatory elements**. *Nat Plants* 2019, **5**:1250-1259.

25. Groen SC, Calic I, Joly-Lopez Z, Platts AE, Choi JY, Natividad M, Dorph K, Mauck WM, 3rd, Bracken B, Cabral CLU, et al.: **The strength and pattern of natural selection on gene expression in rice**. *Nature* 2020, **578**:572-576.
26. Kulkarni SR, Jones DM, Vandepoele K: **Enhanced Maps of Transcription Factor Binding Sites Improve Regulatory Networks Learned from Accessible Chromatin Data**. *Plant Physiol* 2019, **181**:412-425.
27. Maher KA, Bajic M, Kajala K, Reynoso M, Pauluzzi G, West DA, Zumstein K, Woodhouse M, Bubb K, Dorrity MW, et al.: **Profiling of Accessible Chromatin Regions across Multiple Plant Species and Cell Types Reveals Common Gene Regulatory Principles and New Control Modules**. *Plant Cell* 2018, **30**:15-36.
28. Shulse CN, Cole BJ, Ciobanu D, Lin J, Yoshinaga Y, Gouran M, Turco GM, Zhu Y, O'Malley RC, Brady SM, et al.: **High-Throughput Single-Cell Transcriptome Profiling of Plant Cell Types**. *Cell Rep* 2019, **27**:2241-2247 e2244.
29. Ryu KH, Huang L, Kang HM, Schiefelbein J: **Single-Cell RNA Sequencing Resolves Molecular Relationships Among Individual Plant Cells**. *Plant Physiol* 2019, **179**:1444-1456.
30. Jean-Baptiste K, McFaline-Figueroa JL, Alexandre CM, Dorrity MW, Saunders L, Bubb KL, Trapnell C, Fields S, Queitsch C, Cuperus JT: **Dynamics of Gene Expression in Single Root Cells of *Arabidopsis thaliana*. *Plant Cell* 2019, **31**:993-1011.

One of the earliest studies to use single cell RNA-seq data to generate cell type-specific GRNs by analyzing the co-occurrence of TF expression and motif enrichment in specific cell types. This study also examined cell type-specific expression changes in response to heat shock, providing a framework for constructing cell type-specific GRNs responsive to environmental changes.

*31. Denyer T, Ma X, Klesen S, Scacchi E, Nieselt K, Timmermans MCP: **Spatiotemporal Developmental Trajectories in the *Arabidopsis* Root Revealed Using High-Throughput Single-Cell RNA Sequencing**. *Dev Cell* 2019, **48**:840-852 e845.

A TF regulatory network of trichoblast cell differentiation processes was inferred by pseudotime analysis of *Arabidopsis* root single cell RNA-seq data.

*32. Varala K, Marshall-Colon A, Cirrone J, Brooks MD, Pasquino AV, Leran S, Mittal S, Rock TM, Edwards MB, Kim GJ, et al.: **Temporal transcriptional logic of dynamic regulatory networks underlying nitrogen signaling and use in plants**. *Proc Natl Acad Sci U S A* 2018, **115**:6494-6499.

This study incorporated temporal features into the nitrogen response GRN in *Arabidopsis* and predicted TF-target interactions between 155 TFs and 608 nitrogen responsive genes.

33. Stahl PL, Salmen F, Vickovic S, Lundmark A, Navarro JF, Magnusson J, Giacomello S, Asp M, Westholm JO, Huss M, et al.: **Visualization and analysis of gene expression in tissue sections by spatial transcriptomics**. *Science* 2016, **353**:78-82.
*34. Giacomello S, Salmen F, Terebieniec BK, Vickovic S, Navarro JF, Alexeyenko A, Reimegard J, McKee LS, Mannapperuma C, Bulone V, et al.: **Spatially resolved transcriptome profiling in model plant species**. *Nat Plants* 2017, **3**:17061.

The first study to generate high resolution spatial transcriptome profiles in plants, resulted in identification of biological pathways in spatially defined *Arabidopsis* inflorescence meristem micro domains.

35. Kawade K, Tsukaya H: **Probing the stochastic property of endoreduplication in cell size determination of *Arabidopsis thaliana* leaf epidermal tissue**. *PLoS One* 2017, **12**:e0185050.
36. Brooks MD, Cirrone J, Pasquino AV, Alvarez JM, Swift J, Mittal S, Juang CL, Varala K, Gutierrez RA, Krouk G, et al.: **Network Walking charts transcriptional dynamics of nitrogen signaling by integrating validated and predicted genome-wide interactions**. *Nat Commun* 2019, **10**:1569.
37. Zhao H, Zhang W, Chen L, Wang L, Marand AP, Wu Y, Jiang J: **Proliferation of Regulatory DNA Elements Derived from Transposable Elements in the Maize Genome**. *Plant Physiol* 2018, **176**:2789-2803.
38. Lin Y, Meng F, Fang C, Zhu B, Jiang J: **Rapid validation of transcriptional enhancers using agrobacterium-mediated transient assay**. *Plant Methods* 2019, **15**:21.
39. Soyk S, Muller NA, Park SJ, Schmalenbach I, Jiang K, Hayama R, Zhang L, Van Eck J, Jimenez-Gomez JM, Lippman ZB: **Variation in the flowering gene SELF PRUNING 5G promotes day-neutrality and early yield in tomato**. *Nat Genet* 2017, **49**:162-168.
40. Kwon CT, Heo J, Lemmon ZH, Capua Y, Hutton SF, Van Eck J, Park SJ, Lippman ZB: **Rapid customization of Solanaceae fruit crops for urban agriculture**. *Nat Biotechnol* 2020, **38**:182-188.
*41. Rodriguez-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 e478.

This study demonstrated the use of genome editing to validate the effect of *cis*-regulatory elements *in planta* by CRISPR/Cas9 mutagenesis of the promoter region of *SlCLV3*.

42. Lowder LG, Zhou J, Zhang Y, Malzahn A, Zhong Z, Hsieh TF, Voytas DF, Qi Y: **Robust Transcriptional Activation in Plants Using Multiplexed CRISPR-Act2.0 and mTALE-Act Systems.** *Mol Plant* 2018, **11**:245-256.
43. Cao J, Spielmann M, Qiu X, Huang X, Ibrahim DM, Hill AJ, Zhang F, Mundlos S, Christiansen L, Steemers FJ, et al.: **The single-cell transcriptional landscape of mammalian organogenesis**. *Nature* 2019, **566**:496-502.
44. Cusanovich DA, Reddington JP, Garfield DA, Daza RM, Aghamirzaie D, Marco-Ferreres R, Pliner HA, Christiansen L, Qiu XJ, Steemers FJ, et al.: **The cis-regulatory dynamics of embryonic development at single-cell resolution**. *Nature* 2018, **555**:538-542.
45. Cusanovich DA, Hill AJ, Aghamirzaie D, Daza RM, Pliner HA, Berletch JB, Filippova GN, Huang XF, Christiansen L, DeWitt WS, et al.: **A Single-Cell Atlas of In Vivo Mammalian Chromatin Accessibility**. *Cell* 2018, **174**:1309-1324.
46. Budnik B, Levy E, Harmange G, Slavov N: **SCoPE-MS: mass spectrometry of single mammalian cells quantifies proteome heterogeneity during cell differentiation**. *Genome Biol* 2018, **19**:161.
47. Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, Mauck WM, 3rd, Hao Y, Stoeckius M, Smibert P, Satija R: **Comprehensive Integration of Single-Cell Data**. *Cell* 2019, **177**:1888-1902 e1821.
48. Welch JD, Kozareva V, Ferreira A, Vanderburg C, Martin C, Macosko EZ: **Single-Cell Multi-omic Integration Compares and Contrasts Features of Brain Cell Identity**. *Cell* 2019, **177**:1873-1887 e1817.
49. Cao JY, Cusanovich DA, Ramani V, Aghamirzaie D, Pliner HA, Hill AJ, Daza RM, McFaline-Figueroa JL, Packer JS, Christiansen L, et al.: **Joint profiling of chromatin accessibility and gene expression in thousands of single cells**. *Science* 2018, **361**:1380-1385.
50. Li G, Liu Y, Zhang Y, Kubo N, Yu M, Fang R, Kellis M, Ren B: **Joint profiling of DNA methylation and chromatin architecture in single cells**. *Nat Methods* 2019, **16**:991-993.

51. Barro-Trastoy D, Carrera E, Banos J, Palau-Rodriguez J, Ruiz-Rivero O, Tornero P, Alonso JM, Lopez-Diaz I, Gomez MD, Perez-Amador MA: **Regulation of ovule initiation by gibberellins and brassinosteroids in tomato and Arabidopsis: two plant species, two molecular mechanisms**. *Plant J* 2020.
52. Libault M, Pingault L, Zogli P, Schiefelbein J: **Plant Systems Biology at the Single-Cell Level**. *Trends Plant Sci* 2017, **22**:949-960.
53. Brubaker DK, Lauffenburger DA: **Translating preclinical models to humans**. *Science* 2020, **367**:742-743.
54. Arnold CD, Gerlach D, Stelzer C, Boryn LM, Rath M, Stark A: **Genome-wide quantitative enhancer activity maps identified by STARR-seq**. *Science* 2013, **339**:1074-1077.
55. 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**. *Genomics Proteomics Bioinformatics* 2019, **17**:140-153.
56. 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.
57. Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ: **Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position**. *Nat Methods* 2013, **10**:1213-1218.
58. Lu Z, Hofmeister BT, Vollmers C, DuBois RM, Schmitz RJ: **Combining ATAC-seq with nuclei sorting for discovery of cis-regulatory regions in plant genomes**. *Nucleic Acids Res* 2017, **45**:e41.
59. Mumbach MR, Rubin AJ, Flynn RA, Dai C, Khavari PA, Greenleaf WJ, Chang HY: **HiChIP: efficient and sensitive analysis of protein-directed genome architecture**. *Nat Methods* 2016, **13**:919-922.
60. Reynoso MA, Pauluzzi GC, Kajala K, Cabanlit S, Velasco J, Bazin J, Deal R, Sinha NR, Brady SM, Bailey-Serres J: **Nuclear Transcriptomes at High Resolution Using Retooled INTACT.** *Plant Physiol* 2018, **176**:270-281.
61. Birnbaum KD: **Power in Numbers: Single-Cell RNA-Seq Strategies to Dissect Complex Tissues.** *Annu Rev Genet* 2018, **52**:203-221.
62. Matsumoto H, Kiryu H, Furusawa C, Ko MS, Ko SB, Gouda N, Hayashi T, Nikaido I: **SCODE: an efficient regulatory network inference algorithm from single-cell RNA-Seq during differentiation.** *Bioinformatics* 2017, **33**:2314-21.
63. Mirowski P, LeCun Y: **Dynamic factor graphs for time series modeling.** In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases.* Edited by Buntine W, Grobelnik M, Mladenić D, Shawe-Taylor J. Springer, Berlin, Heidelberg; 2009:128-143.

Figure 1

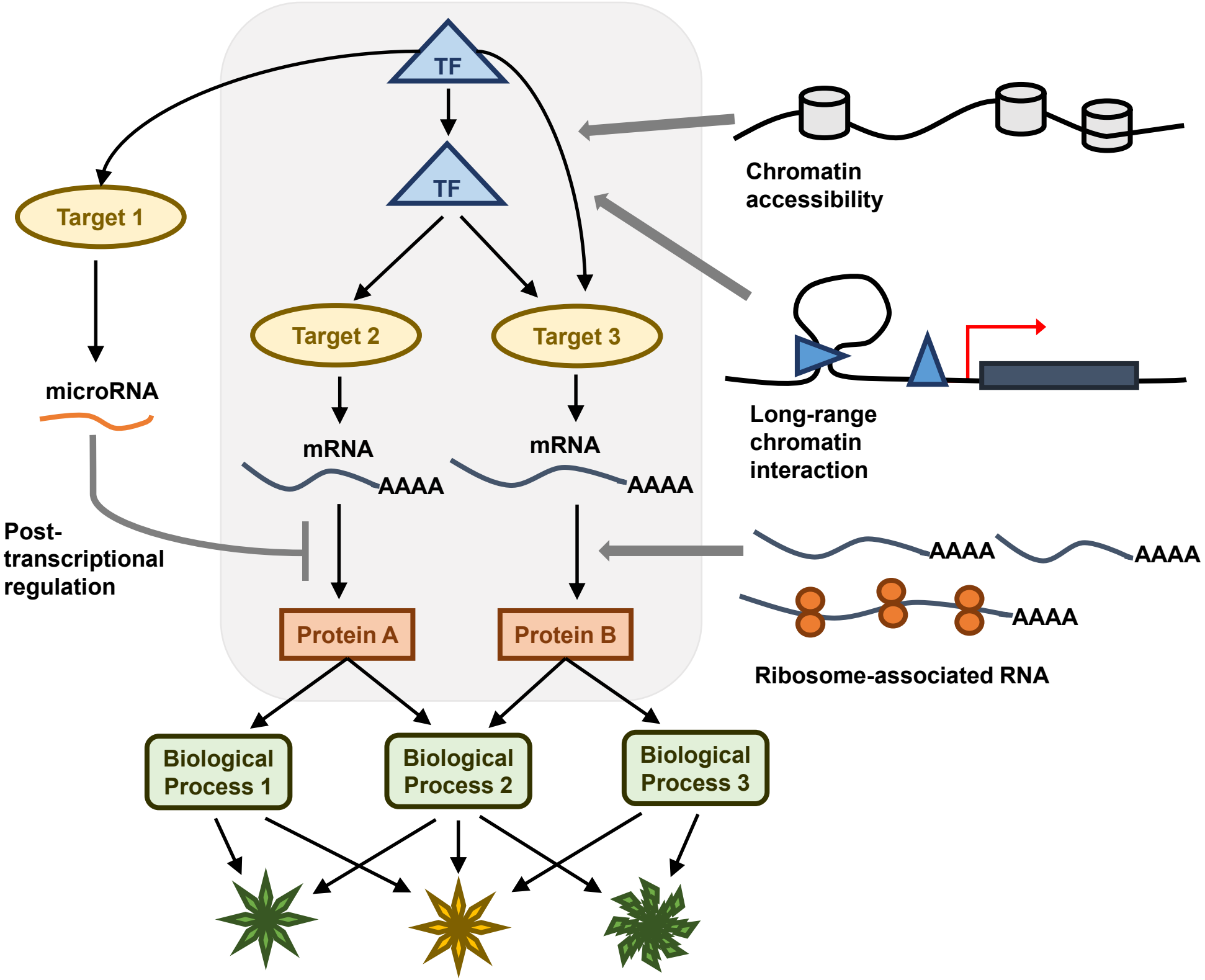
TF
TF
Target 1
Target 2
Target 3
microRNA
mRNA
mRNA
AAAA
AAAA
Post-
transcriptional
regulation
Protein A
Protein B
Chromatin
accessibility
Long-range
chromatin
interaction
AAAA
AAAA
AAAA
Ribosome-associated RNA
Biological
Process 1
Biological
Process 2
Biological
Process 3

Figure 1. Path from GRN to phenotypes

The path from genomic information to phenotypic traits is regulated at multiple levels. A GRN describes the collection of molecules and interactions underlying the biological processes that impact the phenotypes of the plants.

Gray area represents canonical GRN with transcription factor (TF) binding and target gene expression as the center of the network. Black arrow depicts information flow from genome to phenotypes. Outside of the gray area are new features incorporated into the GRN. Dark grey arrow indicates the steps at which these features affect the GRN.

Table

Table 1. Experimental techniques for identification and validation of GRN.

Method	Description	GRN feature identified	References
STARR-Seq (self-transcribing active regulatory region sequencing)	Transfect cells with a construct that contains a minimum promoter upstream of a sequence of interest, followed by sequencing of cellular RNA	Enhancer elements	[21**,54,55]
DAP-seq (DNA affinity purification sequencing)	Affinity purification of genomic DNA by *in vitro* expressed TF followed by sequencing	Transcription factor binding sites	[21**,56]
ATAC-seq (assay for transposase-accessible chromatin sequencing)	Digestion of chromatin at by Tn5 transposase, followed by sequencing	Accessible chromatin regions	[5**,21**,23,27,57,58]
HiChIP	Chromatin immunoprecipitation (ChIP) of factor-directed chromatin contacts, followed by sequencing	Protein mediated chromatin interaction	[21**,59]
TRAP-seq (translating ribosome affinity purification followed by RNA-seq)	Affinity purification of RNA bound by an epitope-tagged ribosomal protein, followed by RNA sequencing	Translating RNA	[5**,22,23]
INTACT (isolation of nuclei tagged in specific cell types)	Transgenic expression of nuclear envelope proteins followed by affinity purification of the tagged nuclei	Nuclei from specific tissue or cell type	[60]
INTACT-ATAC-seq	ATAC-seq with INTACT isolated nuclei	Cell type-specific accessible chromatin regions	[5**,20*,23,27]
TARGET (Transient assay reporting genome-wide effects of transcription factors)	Transient expression of TF in protoplasts followed by RNA sequencing	Target genes activated by a TF	[36]