Characterization of gene regulatory networks in plants using new methods and data types

Klaas Vandepoele[1,2,3] and Kerstin Kaufmann[4]

[1] Ghent University, Department of Plant Biotechnology and Bioinformatics, Technologiepark 71, 9052 Ghent, Belgium

[2] VIB Center for Plant Systems Biology, Technologiepark 71, 9052 Ghent, Belgium

[3] Bioinformatics Institute Ghent, Ghent University, Technologiepark 71, 9052 Ghent, Belgium

[4] Institute of Biology, Humboldt-Universitaet zu Berlin, Berlin, Germany

Key Words

Transcription factor, gene regulatory network, gene activity, genome-wide profiling, plant

Abstract

A major question in plant biology is to understand how plant growth, development and environmental responses are controlled and coordinated by the activities of regulatory factors. Gene regulatory network (GRN) analyses require integrated approaches that combine experimental approaches with computational analyses. A wide range of experimental approaches and tools are now available, such as targeted perturbation of gene activities, quantitative and cell-type specific measurements of dynamic gene activities and systematic analysis of the molecular 'hard-wiring' of the systems. At the computational level, different tools and databases are available to study regulatory sequences, including intuitive visualizations to explore data-driven gene regulatory networks in different plant species. Furthermore, advanced data integration approaches have recently been developed to efficiently leverage complementary regulatory data types and learn context-specific networks.

1. Introduction

Transcriptional regulation is one of the fundamental processes controlling gene expression. Within multicellular organisms, including plants, detailed regulatory programs orchestrate gene activities resulting in phenotypic diversity. This diversity covers the development of specific cell types and organs as well as specific cellular responses to various external stimuli. Transcription factors (TFs), together with other regulatory factors such as miRNAs and other classes of ncRNAs, are key actors influencing gene expression. TFs regulate their target genes by recognizing short sequences in the DNA called TF binding sites (TFBSs). While TF binding can result in the activation or repression of the associated target gene, the interplay of multiple regulators results in the combinatorial control determining spatiotemporal gene expression ***(1, 2)***. The full set of regulatory interactions between a set of functionally related TFs and their target genes form a gene regulatory network (GRN), and these networks are of major interest to get an overview of the organization and complexity of transcriptional regulation at the organismal level. GRNs are pivotal to understanding how different biological processes like growth, development or stress responses are controlled in plants. Examples of well-

characterized signaling cascades cover GRNs controlling root, leaf and flower development, floral transition, photomorphogenesis, as well as cellular responses to plant hormones, such as ethylene and jasmonic acid ***(3-7)***. Furthermore, numerous TFs involved in the transcriptional response to various biotic or abiotic stresses have been reported ***(8-10)***.

Transcription factors can be classified into different TF families based on the presence of protein domains with documented specific DNA binding activity associated with transcriptional regulation. While most TF families are shared between Viridiplantae species (green algae and/or land plants) and other eukaryotes, a subset of families is specific to Viridiplantae or land plants ***(11)***. Plant-specific TF families play key roles in the regulation of various plant developmental, signaling pathways, stress responses, and the biosynthesis of metabolites that, in some cases, are of major importance for plant adaptation ***(12)***. Between 35 and 58 TF families can be found in different plant lineages, each having specific DNA binding signature domains. The number of TFs varies greatly between plant species, from ~100 TFs in the unicellular green alga *Ostreococcus tauri* to >3600 TFs in polyploid species like *Triticum aestivum* (wheat) ***(13)***. Comparison of plant and algal genomes revealed that several TF families have expanded in land plants, including bHLH, NAC, GRAS, AP2/ERF, ASL/LBD, and WRKYs ***(11)***. Whole-genome duplication is the main mechanism responsible for the global expansion of TF gene families. Apart from TF gene copy number variation, the level of divergence, either at the protein level or in spatiotemporal expression, is another major determinant contributing to network complexity, evolution and rewiring of GRNs in different plant species ***(14)***. An assessment of TF DNA binding domain variation found that TF sequence preference divergence varies between TF families. While MYB, C2H2 zinc-finger, and B3 TFs showed the greatest intrafamily divergence, AP2, WRKY, and bHLH were some of the families showing least divergence ***(15)***.

Recent examples of systematic GRN reconstruction in Arabidopsis root and flower development ***(16-20)***, maize leaf development ***(6)***, or in wheat ***(21, 22)*** typically combine dynamic genome-wide expression analyses with information from genome-wide TF binding, e.g. as determined by chromatin immunoprecipitation followed by deep sequencing (ChIP-

seq), DNA affinity purification followed by deep sequencing (DAP-seq) or other ways of mapping of TF binding sites.

Reconstruction of developmental GRNs by combining genome-wide binding and transcriptome analyses revealed a high level of feedback control and cross-regulation between TFs integrating hormonal control, organ and cell-type identity and growth ***(23)***. GRN modelling is potentially complicated by the factor that TFs may act as repressors or activators depending not only on the cis-regulatory grammar ***(24-26)*** but also on TF concentration ***(24)***, and on potentially tissue-type or condition-specific interactions with transcriptional cofactors ***(27)***. A further challenge is to integrate spatial gradients and movement of TFs or other regulatory molecules in GRN modelling, since intercellular communication is an important factor driving developmental patterning and environmental responses, e.g. to light or pathogen attack. Challenges on spatial heterogeneity and cell-type specific regulatory interactions can now be overcome applying techniques for cell type and single-cell omics analysis ***(20)***, which can be complemented by computational reconstruction of cell position in a tissue context ***(28, 29)***. High-resolution transcriptome analyses can be assisted by more sensitive methods measuring transient TF-DNA interactions via DamID-seq or cell-type specific interactions by Cut&Run/Cut&Tag. As the next major milestone, spatial quantitative targeted and untargeted analyses of gene expression and regulatory activities have recently started to be developed. Complementary to this, innovative approaches for genetic perturbation and synthetic regulatory systems have been and are being developed, including the application of dCAS9 variants for targeted induction/repression, cell-type specific knockout by CRISPR/Cas9 or synthetic TFs and TF circuits.

2. Experimental methods to map GRNs in plants

How TFs act in a multicellular context to control cell-type and condition-specific gene expression is still far from understood. Experimental methods that specifically perturb and sensitively analyze TF functions in an in vivo cellular context are needed to address this challenge.

Innovative synthetic systems for cell-type specific activation or knockdown of gene activities have been established. One of these systems is based on a synthetic LhGR TF that combines the DNA-binding domain of a high-affinity DNA-binding mutant of the bacterial *lac* repressor, a transcriptional activation domain of the yeast GAL4 TF and the ligand binding domain of the rat glucocorticoid receptor (GR). Inducible transgene expression is mediated by a synthetic promoter that harbors lacI binding sites (pOp). The system was originally established in plants by the Moore lab ***(30)*** and has been successfully utilized in a number of studies (e.g. ***(31, 32)***). A critical aspect of the cell-type specific induction is to monitor spatiotemporal specificity of the system, which has been addressed by utilizing reporter gene expression, such as GUS ***(32)*** or fluorescent reporters ***(33)*** **(see chapter 2)**. More recently, inducible systems based on other synthetic TFs in combination or recombinase activities have been introduced, providing versatility in tissue-specific activation of transgenes and building synthetic regulatory circuits ***(34, 35)***. In order to modulate the activities of endogenous genes in a specific manner, an endonuclease-deficient version of Cas9 (dCAS9) has been fused with potent transcriptional activation or repression domains (***(36, 37)***, **see chapter 3**). Similar systems have been introduced using synthetic TFs whose DNA-binding specificity can be modulated to target specific endogenous gene loci, such as TALEN or Zinc finger TFs ***(38)***.

Besides specific activation or repression of regulatory proteins in cells and tissues, methods for obtaining context-dependent transcriptional readout and epigenetic status have been developed, facilitated by fluorescent activated cell sorting **(see chapter 5)** or INTACT-based techniques ***(39)***. Single cell genomics approaches are now in place to establish cellular heterogeneity of regulatory protein functions and networks at unprecedented resolution **(see chapter 4)**. Beyond the levels of transcriptional regulation, e.g. via measuring transcript abundance or chromatin status, dynamic changes in proteome and metabolome can be informative readout of developmental or environmental-response time series data **(chapter 6)** to elucidate downstream changes of gene-regulatory programs or possible regulatory mechanisms beyond transcriptional control.

Elucidating the molecular 'hard-wiring', such as protein-DNA interactions and protein-protein interactions, is important towards understanding the mechanistic basis of gene-regulatory interactions. Chromatin immunoprecipitation followed by deep sequencing (ChIP-seq) is a classical method to identify TF-bound genomic regions ***(2, 23, 40)***. In recent years, novel methods to capture not only stable, but also transient or cell-type specific protein-DNA interactions in vivo at genome-wide scale have been established. For example, DNA adenine methyltransferase identification followed by sequencing (DamID-seq ***(41, 42)*****, chapter 7**) makes use of detection of adenine-methylated DNA regions in eukaryotic cells that lack adenine methylation. DamID-seq makes use of a fusion protein of a TF or other DNA-binding protein of interest with E. coli DNA adenine methyltransferase (Dam). This enzyme can methylate the adenine base in GATC sequences that are in spatial proximity of the bound genomic regions. Advantages of the method include that relatively low amounts of tissue are required, and that also transient protein-DNA interactions are captured.

Another approach towards sensitive in vivo mapping of TF-DNA interactions relies on specific cleavage of DNA close to protein-of-interest bound genomic regions. In case of CUT&RUN (Cleavage Under Targets and Release Using Nuclease), protein A/G fused to micrococcal nuclease (MNase) is used to direct nuclease activity towards genomic regions bound by a specific protein, facilitated by interaction of protein A/G with an antibody against the protein-of interest ***(39)***. CUT&TAG (Cleavage Under Targets and Tagmentation) is a related approach (***(43, 44)***, **chapter 8**) , in which the protein A/G is fused to hyperactive Tn5 transposase pre-loaded with sequencing adaptors, allowing for simultaneous cleavage of DNA and library preparation. Both methods are highly sensitive and applicable to low amounts of tissue, yet proper controls to correct for non-specific cuts in accessible genomic regions. Once TF-bound genomic regions have been identified, a final challenge is to associate the regulatory regions to which the TF binds with potentially directly regulated target genes. This is particularly the case in larger genomes with complex and distantly located regulatory regions. Chromatin Interaction Analysis by Paired-End Tag Sequencing (ChIA-PET) addresses this question by determining protein-of-interest-mediated long-range chromatin interactions (***(45)***, **chapter 12**).

The method essentially combines ChIP with chromatin conformation capture (3C) and has been successfully used in species such as rice and maize ***(45)***. Related methods can also be applied to capture higher-order chromatin interactions in targeted or untargeted manner ***(45)***.

A limitation to in vivo methods for identifying protein-DNA interactions is that they typically do not allow to elucidate the precise mechanisms of TF DNA-binding, e.g. the requirement for combinatorial protein-protein interactions or DNA recognition. Mechanistic knowledge is however crucial for generating predictive knowledge on conditional or cell-type specific regulatory interactions. In vitro methods to elucidate physical parameters of DNA-binding, such as specificity and affinity, can be applied towards this goal. DNA affinity purification followed by sequencing (DAP-seq) is a method that relies on isolating TF-bound DNA fragments that represent sheared genomic DNA ***(46)***. Affinity purified DNA sequences are subjected to high-throughput sequencing and mapped to the genome. While DAP-seq typically identifies DNA regions bound by a specific TF that is produced in vitro or in *E. coli*, a modified version of the protocol can also be used to identify DNA-binding sites of heteromeric TF complexes (***(47)***, **chapter 9**). While DAP-seq can provide an overall view on potential DNA binding sites of a TF, SELEX-seq (Systematic Evolution of Ligands by Exponential Enrichment followed by sequencing) can be alternatively used to characterize and compare TF DNA-binding specificities (***(48)***, **chapter 10**). Instead of starting with a pool of genomic DNA, SELEX-seq uses short (~20-40 bp) randomized DNA pools as input for the TF-DNA-binding reaction. Protein-DNA complexes are affinity purified and the purified DNA is used as input for additional rounds of enrichment, and DNA from several rounds of enrichment is sequenced, followed by computational analysis of enriched sequence patterns. Both DAP-seq and SELEX-seq have been successfully used to characterize DNA-binding sites and DNA-binding specificities of plant transcription factors.

Besides protein-DNA interactions, the ‘hard-wiring’ of GRNs is also influenced by protein-protein interactions, such as combinatorial TF interactions as well as interactions of TFs and cofactors. A large number of approaches to probe physical interactions of TFs and other proteins in vitro or in planta have been established. Typical targeted *in planta* methods to study

interactions of two specific proteins include bimolecular fluorescence complementation (BiFC), split luciferase (SLC) or Förster Resonance Energy Transfer combined with Fluorescence Lifetime Imaging (FRET-FLIM). FRET-FLIM is particularly suitable to identify cell-type specific interactions of regulatory proteins ***(49)***. All methods require appropriate control experiments in the experimental design (see, e.g. ***(50)***).

Untargeted *in vivo* methods for identifying regulatory protein interactions are mostly based on mass spectrometry-based identification of proteins. 'Classical' methods are usually based affinity purification of proteins-of-interest fused to a tag. Such tags include simple tags like GREEN FLUORESCENT PROTEIN (GFP, ***(51)***, **chapter 11**) or tags for tandem affinity purification ***(52)***. More recently, proximity labelling has been introduced to identify complex partners of low-abundant regulatory proteins ***(53)***. The method has also been adapted to identify the nuclear proteome of a rare cell type ***(53)***. Essentially, modified versions of bacterial biotin ligase (BirA), such as TurboID ***(54)***, are fused to a protein of interest and expressed in plants. Upon exogenous supply of the substrate biotin, BirA binds and modifies biotin which can then covalently bind to lysine residues in proximally located proteins, such as complex partners of the protein of interest. For all the untargeted methods, careful experimental design and quantitative data analysis is required to identify protein complex partners.

3. Data analysis and inference methods to accurately model gene regulation

Given the wealth of experimental profiling methods to characterize gene regulation in plants, new analysis and integration methods are needed to fully exploit the complementary information present in the obtained datasets. Experimental methods available to study gene regulation capture biological insights at different levels, ranging from the genomic DNA, over TF proteins, to regulatory interactions between regulators and target genes. Context-specific information, either from different organs, tissues or cell types, together with temporal dynamics, offer another level of information to comprehensively and accurately map plant GRNs in vivo. While each experimental method requires dedicated data analysis, standardization and

automation are of utmost importance to guarantee reproducibility. Furthermore, access to datasets processed using the same underlying methods is essential when i) comparing results from different organs/cell types, species or research labs and ii) devising integration strategies aiming to combine the complementary knowledge captured by different data types **(*(55-57)*, chapter 13, chapter 14)**.

Several studies have demonstrated how the integration of complementary omics data types offers new insights on biological networks ***(58)***. While a targeted or context-specific setup focuses on a specific developmental process or environmental response of interest, untargeted approaches have the potential to study gene regulation covering a broad array of processes, often making use of publicly available datasets. Zander and co-workers applied a targeted multi-omics network approach, where the dynamic profiling of regulatory interactions, chromatin state, transcriptome and (phospho)proteome was used to delineate jasmonate signaling networks in Arabidopsis ***((3), chapter 15)***. Using an untargeted approach, the integrative analysis of a transcriptomics and interactomics data was used to infer functional and regulatory annotations for >5,000 unknown Arabidopsis genes ***(59)***. While condition-specific co-expression modules together with guilt-by-association were first used to predict new gene functions, the integration of experimental protein-DNA and protein-protein interaction networks added physical and/or regulatory support.

Various databases and platforms offer data-driven approaches to study gene regulation and infer GRNs in plants, including PlantRegMap, AtRegNet, AIV2, ConnecTF, and TF2Network ***((13, 55, 56, 60, 61), chapter 13, chapter 17, chapter 19, chapter 20)***. These resources, as well as related initiatives, leverage diverse experimental data types, including expression data, chromatin data, TF motifs, genomic variation data ***(Chapter 18)***, protein-DNA and protein-protein interactions, to identify regulatory interactions. Querying these regulatory data types for a specific TF, gene, pathway, or biological process of interest, immediately sheds light on the available experimental evidence and allows to formulate new research questions or hypotheses. These complementary experimental datasets also form the basis for powerful integration strategies, for example using machine learning, to infer and prioritize new

regulatory interactions ***((62, 63), chapter 21)***. ConSReg is a condition specific regulatory network inference engine which uses a machine learning approach to integrate expression data, TF–DNA binding data and open chromatin data ***((64), chapter 16)***. It focuses on using protein-DNA interaction data and open chromatin data to predict the combinations of TFs that can best explain observed differential gene expression under different environmental perturbations or cell types. In a related approach ***(63)***, different input networks capturing complementary information about DNA motifs, open chromatin, TF-binding and expression-based regulatory interactions were combined using a supervised learning approach, resulting in an integrated Arabidopsis gene regulatory network (iGRN) covering 1.7 million interactions. This iGRN has a similar performance to recover functional interactions compared to state-of-the-art experimental methods and allows to infer known and novel TF functions, as demonstrated for regulators predicted to be involved in reactive oxygen species stress regulation ***(63)***.

Apart from experimental data, meta-data information about genotypes (e.g. mutant or over-expression lines) and ecotypes, growth conditions, sampling procedure, stress application or perturbation, and technical properties of the applied profiling method is equally important to identify, and potentially correct, biological or technical biases leading to confounders during the analysis of regulatory datasets. As an example, the integration of different epigenomic datasets, such as accessible chromatin regions, various histone modifications, or DNA methylation, only makes sense if samples derived from the same organ and developmental stage or combined. As such, databases implementing sample validation and certification services improve the FAIRness (i.e. level of findability, accessibility, interoperability, and reusability) of samples at the source (i.e. at submission time) and facilitate extracting novel biological knowledge through the integrative analysis of different datasets. An example of such an integrative approach was the identification of unmethylated regions (UMRs) which provide useful information for identification of functional genes and CREs ***(65)***. The comparison of UMRs in different plant genomes with publicly available tissue-specific chromatin accessibility and gene expression information provided evidence for the functional role of these regions.

For plant species with large genomes, such as maize and barley, the authors suggested that the identification of UMRs from a single tissue can assist in delineating a fairly complete catalog of potential regulatory elements and expressed genes across many developmental stages.

4. Conclusions

In summary, the development and application of novel experimental and computational methodologies allows plant scientists to look with unprecedented resolution to gene activities, further improving our understanding of plant GRNs. The complementarity of these methods is a blessing and curse at the same time: while they allow to measure specific molecular activities and generate novel insights for specific TFs, promoters or target genes of interest, the integration of these different layers of information, measuring regulatory properties at different scales, also brings new challenges to unify all this knowledge in comprehensive and understandable biological models.

Acknowledgements

The authors wish to thank the German Research Foundation (DFG) [458750707, 438774542, 270050988, 355312821 to KK], the Research Foundation-Flanders (FWO) for ELIXIR Belgium [I002819N to K.V] and Ghent University BOF Grant [BOF24Y2019001901 to K.V.] for support. We apologize to all authors whose work could not be cited due to space constraints.

References

1. Schmitz, R.J., E. Grotewold, and M. Stam (2022) Cis-regulatory sequences in plants: Their importance, discovery, and future challenges. Plant Cell 34:718-741
2. Heyndrickx, K.S., et al. (2014) A Functional and Evolutionary Perspective on Transcription Factor Binding in Arabidopsis thaliana. Plant Cell 26:3894-910
3. Zander, M., et al. (2020) Integrated multi-omics framework of the plant response to jasmonic acid. Nat Plants 6:290-302
4. Chang, K.N., et al. (2013) Temporal transcriptional response to ethylene gas drives growth hormone cross-regulation in Arabidopsis. Elife 2:e00675
5. Lorenzo, O. and R. Solano (2005) Molecular players regulating the jasmonate signalling network. Curr Opin Plant Biol 8:532-40
6. Tu, X., et al. (2020) Reconstructing the maize leaf regulatory network using ChIP-seq data of 104 transcription factors. Nat Commun 11:5089

7. Sullivan, A.M., et al. (2014) Mapping and dynamics of regulatory DNA and transcription factor networks in A. thaliana. Cell Rep 8:2015-2030
8. Walley, J.W. and K. Dehesh (2010) Molecular mechanisms regulating rapid stress signaling networks in Arabidopsis. J Integr Plant Biol 52:354-9
9. Urano, K., et al. (2010) 'Omics' analyses of regulatory networks in plant abiotic stress responses. Curr Opin Plant Biol 13:132-8
10. Ma, S. and H.J. Bohnert (2007) Integration of Arabidopsis thaliana stress-related transcript profiles, promoter structures, and cell-specific expression. Genome Biol 8:R49
11. Lehti-Shiu, M.D., et al. (2017) Diversity, expansion, and evolutionary novelty of plant DNA-binding transcription factor families. Biochim Biophys Acta Gene Regul Mech 1860:3-20
12. Yamasaki, K., et al. (2013) DNA-binding domains of plant-specific transcription factors: structure, function, and evolution. Trends Plant Sci 18:267-76
13. Jin, J., et al. (2017) PlantTFDB 4.0: toward a central hub for transcription factors and regulatory interactions in plants. Nucleic Acids Res 45:D1040-D1045
14. Jones, D.M. and K. Vandepoele (2020) Identification and evolution of gene regulatory networks: insights from comparative studies in plants. Curr Opin Plant Biol 54:42-48
15. Lambert, S.A., et al. (2019) Similarity regression predicts evolution of transcription factor sequence specificity. Nat Genet 51:981-989
16. Brady, S.M., et al. (2011) A stele-enriched gene regulatory network in the Arabidopsis root. Mol Syst Biol 7:459
17. Chen, D., et al. (2018) Architecture of gene regulatory networks controlling flower development in Arabidopsis thaliana. Nat Commun 9:4534
18. Moreno-Risueno, M.A., et al. (2015) Transcriptional control of tissue formation throughout root development. Science 350:426-30
19. O'Maoileidigh, D.S., E. Graciet, and F. Wellmer (2014) Gene networks controlling Arabidopsis thaliana flower development. New Phytol 201:16-30
20. Ferrari, C., N. Manosalva Perez, and K. Vandepoele (2022) MINI-EX: Integrative inference of single-cell gene regulatory networks in plants. Mol Plant 15:1807-1824
21. Zhang, Y., et al. (2021) Evolutionary rewiring of the wheat transcriptional regulatory network by lineage-specific transposable elements. Genome Res 31:2276-89
22. Zhang, Y., et al. (2022) Transposable elements orchestrate subgenome-convergent and -divergent transcription in common wheat. Nat Commun 13:6940
23. Gaudinier, A. and S.M. Brady (2016) Mapping Transcriptional Networks in Plants: Data-Driven Discovery of Novel Biological Mechanisms. Annu Rev Plant Biol 67:575-94
24. Perales, M., et al. (2016) Threshold-dependent transcriptional discrimination underlies stem cell homeostasis. Proc Natl Acad Sci U S A 113:E6298-E6306
25. Rodriguez, K., et al. (2022) Concentration-dependent transcriptional switching through a collective action of cis-elements. Sci Adv 8:eabo6157
26. White, M.A., et al. (2016) A Simple Grammar Defines Activating and Repressing cis-Regulatory Elements in Photoreceptors. Cell Rep 17:1247-1254
27. Plant, A.R., A. Larrieu, and B. Causier (2021) Repressor for hire! The vital roles of TOPLESS-mediated transcriptional repression in plants. New Phytol 231:963-973
28. Laureyns, R., et al. (2022) An in situ sequencing approach maps PLASTOCHRON1 at the boundary between indeterminate and determinate cells. Plant Physiol 188:782-794
29. Liu, C., et al. (2022) A spatiotemporal atlas of organogenesis in the development of orchid flowers. Nucleic Acids Res 50:9724-9737
30. Samalova, M., B. Brzobohaty, and I. Moore (2005) pOp6/LhGR: a stringently regulated and highly responsive dexamethasone-inducible gene expression system for tobacco. Plant J 41:919-35
31. Lopez-Salmeron, V., et al. (2019) Inducible, Cell Type-Specific Expression in Arabidopsis thaliana Through LhGR-Mediated Trans-Activation. J Vis Exp
32. Samalova, M., C. Kirchhelle, and I. Moore (2019) Universal Methods for Transgene Induction Using the Dexamethasone-Inducible Transcription Activation System pOp6/LhGR in Arabidopsis and Other Plant Species. Curr Protoc Plant Biol 4:e20089

33. Schurholz, A.K., et al. (2018) A Comprehensive Toolkit for Inducible, Cell Type-Specific Gene Expression in Arabidopsis. Plant Physiol 178:40-53
34. Brophy, J.A.N., et al. (2022) Synthetic genetic circuits as a means of reprogramming plant roots. Science 377:747-751
35. Lloyd, J.P.B., et al. (2022) Synthetic memory circuits for stable cell reprogramming in plants. Nat Biotechnol 40:1862-1872
36. Lowder, L.G., A. Malzahn, and Y. Qi (2018) Plant Gene Regulation Using Multiplex CRISPR-dCas9 Artificial Transcription Factors. Methods Mol Biol 1676:197-214
37. Pan, C., et al. (2021) CRISPR-Act3.0 for highly efficient multiplexed gene activation in plants. Nat Plants 7:942-953
38. Liu, W. and C.N. Stewart, Jr. (2016) Plant synthetic promoters and transcription factors. Curr Opin Biotechnol 37:36-44
39. Deal, R.B. and S. Henikoff (2010) A simple method for gene expression and chromatin profiling of individual cell types within a tissue. Dev Cell 18:1030-40
40. van Mourik, H., et al. (2015) Characterization of in vivo DNA-binding events of plant transcription factors by ChIP-seq: experimental protocol and computational analysis. Methods Mol Biol 1284:93-121
41. Alvarez, J.M., et al. (2020) Transient genome-wide interactions of the master transcription factor NLP7 initiate a rapid nitrogen-response cascade. Nat Commun 11:1157
42. Wu, F., B.G. Olson, and J. Yao (2016) DamID-seq: Genome-wide Mapping of Protein-DNA Interactions by High Throughput Sequencing of Adenine-methylated DNA Fragments. J Vis Exp e53620
43. Kaya-Okur, H.S., et al. (2019) CUT&Tag for efficient epigenomic profiling of small samples and single cells. Nat Commun 10:1930
44. Ouyang, W., et al. (2021) Rapid and Low-Input Profiling of Histone Marks in Plants Using Nucleus CUT&Tag. Front Plant Sci 12:634679
45. Ouyang, W., et al. (2020) Unraveling the 3D Genome Architecture in Plants: Present and Future. Mol Plant 13:1676-1693
46. O'Malley, R.C., et al. (2016) Cistrome and Epicistrome Features Shape the Regulatory DNA Landscape. Cell 165:1280-1292
47. Lai, X., et al. (2020) Genome-wide binding of SEPALLATA3 and AGAMOUS complexes determined by sequential DNA-affinity purification sequencing. Nucleic Acids Res 48:9637-9648
48. Slattery, M., et al. (2011) Cofactor binding evokes latent differences in DNA binding specificity between Hox proteins. Cell 147:1270-82
49. Long, Y., et al. (2017) In vivo FRET-FLIM reveals cell-type-specific protein interactions in Arabidopsis roots. Nature 548:97-102
50. Strotmann, V.I. and Y. Stahl (2022) Visualization of in vivo protein-protein interactions in plants. J Exp Bot 73:3866-3880
51. Smaczniak, C., et al. (2012) Characterization of MADS-domain transcription factor complexes in Arabidopsis flower development. Proc Natl Acad Sci U S A 109:1560-5
52. Van Leene, J., et al. (2015) An improved toolbox to unravel the plant cellular machinery by tandem affinity purification of Arabidopsis protein complexes. Nat Protoc 10:169-87
53. Mair, A., et al. (2019) Proximity labeling of protein complexes and cell-type-specific organellar proteomes in Arabidopsis enabled by TurboID. Elife 8:
54. Branon, T.C., et al. (2018) Efficient proximity labeling in living cells and organisms with TurboID. Nat Biotechnol 36:880-887
55. Brooks, M.D., et al. (2021) ConnecTF: A platform to integrate transcription factor-gene interactions and validate regulatory networks. Plant Physiol 185:49-66
56. Kulkarni, S.R., et al. (2018) TF2Network: predicting transcription factor regulators and gene regulatory networks in Arabidopsis using publicly available binding site information. Nucleic Acids Res 46:e31
57. Fu, L.Y., et al. (2022) ChIP-Hub provides an integrative platform for exploring plant regulome. Nat Commun 13:3413

58. Depuydt, T., B. De Rybel, and K. Vandepoele (2022) Charting plant gene functions in the multi-omics and single-cell era. Trends Plant Sci
59. Depuydt, T. and K. Vandepoele (2021) Multi-omics network-based functional annotation of unknown Arabidopsis genes. Plant J
60. Palaniswamy, S.K., et al. (2006) AGRIS and AtRegNet. a platform to link cis-regulatory elements and transcription factors into regulatory networks. Plant Physiol 140:818-29
61. Dong, S., et al. (2019) Proteome-wide, Structure-Based Prediction of Protein-Protein Interactions/New Molecular Interactions Viewer. Plant Physiol 179:1893-1907
62. Meng, X., et al. (2021) Predicting transcriptional responses to cold stress across plant species. Proc Natl Acad Sci U S A 118:
63. De Clercq, I., et al. (2021) Integrative inference of transcriptional networks in Arabidopsis yields novel ROS signalling regulators. Nature Plants 7:500-513
64. Song, Q., et al. (2020) Prediction of condition-specific regulatory genes using machine learning. Nucleic Acids Res 48:e62
65. Crisp, P.A., et al. (2020) Stable unmethylated DNA demarcates expressed genes and their cis-regulatory space in plant genomes. Proc Natl Acad Sci U S A 117:23991-24000