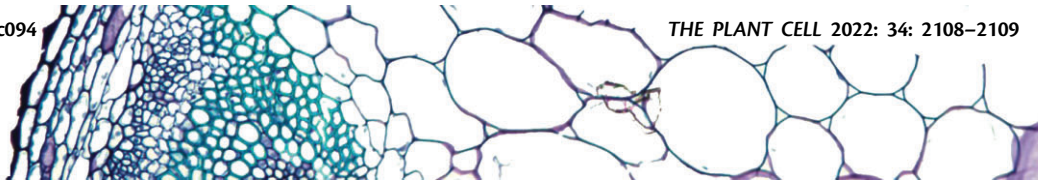
https://doi.org/10.1093/plcell/koac094 THE PLANT CELL 2022: 34: 2108–2109



Decoding the cis-regulation of tomato fruit development with deep learning

Humberto Herrera-Ubaldo [1,2,*]

1 Assistant Features Editor, *The Plant Cell*, American Society of Plant Biologists, USA
2 Unidad de Genómica Avanzada (UGA-Langebio), Centro de Investigación y de Estudios Avanzados (CINVESTAV-IPN), Irapuato 36824, Mexico

*Author for correspondence: humberto.herrera@cinvestav.mx

In Brief

The transcriptional regulation underlying biological processes is mostly controlled by transcription factors (TFs) and co-factors (*trans*-factors) that bind specific sequences in the promoter regions (*cis*-elements) to activate or repress gene expression (see Figure). Natural variation in promoter regions contributes to the diversity of expression patterns seen in nature. Modifications within regulatory regions can also generate novel alleles that alter crop traits; for example, CRISPR/Cas9-induced mutations in the promoters of key regulators of meristematic activity led to enhanced inflorescence branching and increased locule number and fruit size in tomato (Rodríguez-Leal et al., 2017).

In this issue of *The Plant Cell*, work by **Takashi Akagi, Kanae Masuda, Eriko Kuwada, and colleagues** (Akagi et al., 2022) advances our understanding of the precise control of gene expression during tomato fruit development. Artificial intelligence methods (including deep learning) allow the analysis of massive amounts of data to find patterns and connections between variables. In plant biology, a recent example of using these algorithms involves the study of gene expression patterns in response to wounding (Moore et al., 2022). Here, Akagi et al. used an "explainable" deep learning framework to identify *cis*-regulatory elements (CREs) that can predict gene expression during tomato fruit ripening.

The first step was to determine sequence patterns underlying TF binding. The authors used the so-called *cistrome* datasets that include data from DNA-affinity purification and sequencing (DAP-seq) (O'Malley et al., 2016), which describe DNA sequences associated with TF binding, and used only high-confidence binding motifs for 370 TFs from this dataset. The deep learning models allowed the authors to identify known motifs that have been extensively biologically characterized, and also motifs with minor sequence variations. One advantage of such variant discovery is that it can be applied to study gene regulation in other related genomes. In this study, the authors looked for TF-binding sites identified from the 370 TFs in the 1-kb promoter regions of the 34,066 genes encoded in the tomato genome to predict potential CREs for each TF.

The next step was to identify key CREs underlying transcriptional changes during tomato ripening (transition from green to red fruits) and to then use these CREs to predict gene expression patterns. The authors used a previously reported dataset of tomato gene expression (Shinozaki et al., 2018), which characterizes gene expression in the pericarp at different ripening stages. With these data, they identified genes significantly upregulated or downregulated during the transition from the mature green to the breaker developmental stage. The classification model associated with significantly upregulated genes during the transition to the breaker stage of tomato fruit development yielded the best prediction accuracy, and therefore CREs were only predicted for this set of genes. The promoters of upregulated genes displayed enrichment in binding sites for NAC, C2H2, MADS-box, G2-like, and ERF TF families. Some of the well-described regulators of tomato fruit ripening (i.e., *NON-RIPENING*, *SlZFP2*, or *RIPENING INHIBITOR*) belong to those families.

The authors chose a subset of genes to verify the relevance of CREs predicted to control their expression. For instance, the promoter region of *Aminocyclopropane-1-carboxylic Acid Synthase 2* (*ACS2*), a key enzyme in the ethylene biosynthesis pathway, contained relevant binding sites for NAC and MADS-box TFs (see Figure). The functional validation of these motifs was determined by analyzing a mutated

Received March 03, 2022. Accepted March 14, 2022. Advance access publication March 17, 2022

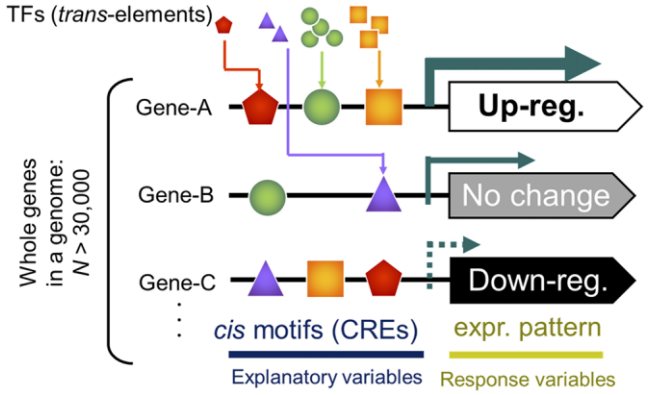


Figure TFs bind regulatory elements in promoter regions (CREs) to regulate gene expression and modify expression levels. In this work, CREs were analyzed to understand expression patterns at the genome-wide level in tomato. Adapted from Akagi et al. (2022), Figure 1.

promoter sequence (*pACS2mut*) and assessing the probability of up-regulation in silico using their reported deep learning algorithms. Then, the same "*pACS2mut*" and wild-type sequences were tested for their ability to bind cognate TFs in vitro using an electrophoretic mobility shift assay and the capacity for transcriptional regulation using two transient reporter assays, with either luciferase or GFP. The NON-RIPENING TF, a putative regulator of *ACS2* gene expression, did not bind pACS2mut; additionally, the expression of *ACS2* was significantly reduced, revealing the contribution of the CREs to the control of gene expression.

The deep learning framework described by Akagi et al. is similar to a novel approach that uses "transformer modules" (Avsec et al., 2021). Akagi et al. used a more limited set of transcriptional data to mine predictive relationships but has the advantage of using a less "species-specific" DAPseq dataset. Further advances in predictive ability could be gained by integrating these new deep learning frameworks and using more comprehensive "omic" datasets. Additionally, some extra layers of regulation could be addressed, such as indirect binding of TFs via protein interactions and modifications at the chromatin level. Integrating these diverse regulatory elements will allow the design of novel alleles with specific expression patterns.

References

Akagi T, Masuda K, Kuwada E, Takeshita K, Kawakatsu T, Ariizumi T, Kubo Y, Ushijima K, Uchida S (2022) Genome-wide cis-decoding for expression design using cistrome data and explainable deep learning. Plant Cell **34:** 2174–2187

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

Moore BM, Lee YS, Wang P, Azodi C, Grotewold E, Shiu S-H (2022) Modeling temporal and hormonal regulation of plant transcriptional response to wounding. Plant Cell **34:** 867–888

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

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

Shinozaki Y, Nicolas P, Fernandez-Pozo N, Ma Q, Evanich DJ, Shi Y, Xu Y, Zheng Y, Snyder SI, Martin LBB, et al. (2018) High-resolution spatiotemporal transcriptome mapping of tomato fruit development and ripening. Nat Commun **9**: 364