









Modeling temporal and hormonal regulation of plant transcriptional response to wounding

Bethany M. Moore ^{1,2,3,*} YunSun Lee ⁴ Peipei Wang ² Christina Azodi ⁵ Erich Grotewold ⁴ and Shin-Han Shiu ^{2,3,6,*,†}

1 Department of Botany, University of Wisconsin-Madison, Madison, Wisconsin 53706, USA

2 Department of Plant Biology, Michigan State University, East Lansing, Michigan 48824, USA

3 Ecology, Evolutionary Biology, and Behavior program, Michigan State University, East Lansing, Michigan 48824, USA

4 Department of Biochemistry and Molecular Biology, Michigan State University, East Lansing, Michigan 48824, USA

5 St. Vincent's Institute of Medical Research, Fitzroy 3065, Victoria, Australia

6 Department of Computational Mathematics, Science and Engineering, Michigan State University, East Lansing, Michigan 48824, USA

*Author for correspondence: bmoore22@wisc.edu (B.M.M.), shius@msu.edu (S.-H.S.).

Joint first authors (B.M.M., Y.S.L.).

†Senior author.

B.M.M. and S.-H.S. designed the research and revised the article. B.M.M. performed most of the computational analyses and wrote the manuscript. Y.S.L. performed the wet-bench experiments and contributed to the writing of the manuscript. C.B.A. contributed new computational tools, interpreted data, and revised the article. P.W. interpreted data and revised the article. E.G. designed the experimental component and revised the article. The authors responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors (<https://academic.oup.com/plcell>) are: Bethany M. Moore (bmoore22@wisc.edu) and Shin-Han Shiu (shius@msu.edu).

Abstract

Plants respond to wounding stress by changing gene expression patterns and inducing the production of hormones including jasmonic acid. This wounding transcriptional response activates specialized metabolism pathways such as the glucosinolate pathways in *Arabidopsis thaliana*. While the regulatory factors and sequences controlling a subset of wound-response genes are known, it remains unclear how wound response is regulated globally. Here, we show these responses are regulated by incorporating putative *cis*-regulatory elements, known transcription factor binding sites, *in vitro* DNA affinity purification sequencing, and DNase I hypersensitive sites to predict genes with different wound-response patterns using machine learning. We observed that regulatory sites and regions of open chromatin differed between genes upregulated at early and late wounding time-points as well as between genes induced by jasmonic acid and those not induced. Expanding on what we currently know, we identified *cis*-elements that improved model predictions of expression clusters over known binding sites. Using a combination of genome editing, *in vitro* DNA-binding assays, and transient expression assays using native and mutated *cis*-regulatory elements, we experimentally validated four of the predicted elements, three of which were not previously known to function in wound-response regulation. Our study provides a global model predictive of wound response and identifies new regulatory sequences important for wounding without requiring prior knowledge of the transcriptional regulators.

Introduction

Plants respond to environmental stresses by reprogramming their pattern of gene expression that triggers chemical and physiological responses (Bostock et al., 2014). These stress

responses can be important to plant survival in their respective niches and are likely subjected to selection (Bostock et al., 2014). Wound stress is a common stress experienced by plants when they are under certain biotic stresses such

Received November 24, 2020. Accepted November 18, 2021. Advance access publication December 3, 2021

© The Author(s) 2021. Published by Oxford University Press on behalf of American Society of Plant Biologists.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs licence (<https://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial reproduction and distribution of the work, in any medium, provided the original work is not altered or transformed in any way, and that the work is properly cited. For commercial use, please contact journals.permissions@oup.com

Open Access

as attack by insects or abiotic stress such as wind damage and can induce a chemical response that produces compounds of human interest (Jacobo-Velázquez et al., 2015). Response to stresses such as wounding requires gene expression reprogramming, a complex process that involves multiple levels of regulation. At the DNA sequence level, short stretches of DNA (*cis*-regulatory elements, CREs) are recognized and bound by transcription factors (TFs) that can activate or repress gene expression (Wittkopp and Kalay, 2012). Beyond the level of DNA sequence, chromatin structure can influence whether a regulatory element is accessible to a TF and can be modified based on stress–response signals (Asensi-Fabado et al., 2017). Finally, reprogramming can also occur by modifying or turning over mRNA (Glisovic et al., 2008; Hutvagner and Simard, 2008). Stress responses change over time, adding temporal complexity to transcriptional response. For example, after an initial response, genes that are turned on may act to turn on or off other genes, resulting in cascading effects. This type of gene expression reprogramming mechanism is beneficial for plants when different responses are needed at different times. For example, response to wounding stress in plants changes over time as the plant first needs to recognize damaging agents, then responds by sending various hormone signals, and ultimately repairs the wound (Ikeuchi et al., 2017). This means that stress-responsive genes may be regulated differently depending on when they need to be expressed.

The production of various hormone signals allows plants to coordinate their response to different stresses because the interactions of certain hormones can regulate a specific response from the plant by changing the expression of certain genes. For example, response to wounding stress involves several hormones, with the most ubiquitous signal being jasmonic acid (JA; Howe and Jander, 2008). After wounding, JA levels increase and bind to Jasmonate ZIM (JAZ) domain repressor proteins, which allows MYC2 TFs and other basic helix–loop–helix (bHLH) TFs to become active (Chung et al., 2008). MYC2 TFs then activate wounding responses, such as JA biosynthesis, to amplify the JA signal and activate other defensive processes (Chung et al., 2008). Additional hormones interact with JA to moderate wounding response. For instance, while JA induces the expression of certain wound-response genes, ethylene simultaneously represses the expression of these genes at the damaged site in order to make sure the correct spatial response pattern is produced (Rojo et al., 1999). Ethylene also works synergistically with JA to fine-tune wounding response by inducing the expression of proteinase inhibitor genes (O'Donnell et al., 1996) and by activating ETHYLENE RESPONSE FACTOR 1, another TF that triggers defense responses (Lorenzo et al., 2003). Absciscic acid (ABA), which is induced in response to many abiotic stresses, is also induced by wounding (León et al., 2001). While ethylene, ABA, and JA rapidly respond to wounding, other hormones such as auxin and cytokinin, start to accumulate around 12 hours (h) after

wounding occurs and are involved in signaling for the regulation of expression of genes that work to repair the wound (Ikeuchi et al., 2017). While a great deal is known about hormone signaling in response to wounding, it is unclear what other regulatory mechanisms are involved in response to wounding and how these mechanisms interact with hormone signals.

Wounding can also induce the production of specialized metabolites that can deter further stress. For example, after wounding stress, *Arabidopsis thaliana* (*Arabidopsis*) activates glucosinolate pathways. The glucosinolates and the bioproducts generated from their degradation affect the plant's interactions with biotic stressors, such as microbes and herbivores (Yan and Chen, 2007). Additionally, mutants with decreased glucosinolate levels show greater susceptibility to the necrotrophic fungus *Fusarium oxysporum* (Tierens et al., 2001). Glucosinolate production is regulated by JA, salicylic acid, and ethylene. These hormones work together to modulate glucosinolate levels in response to stress, by activating MYB (myeloblastosis) and DNA-binding with one finger TFs (Yan and Chen, 2007). Additionally, glucosinolates can be divided into different types, such as indole or aliphatic glucosinolates, and the production of these may be induced by various stresses and regulated in different ways (Yan and Chen, 2007). While specific TFs have been shown to turn on glucosinolate biosynthesis (Frerigmann and Gigolashvili, 2014), the regulatory elements or chromatin structure of how and when these TFs bind has not been resolved.

At the *cis*-regulatory level, a few CREs underlying wound-response regulation have been discovered experimentally. An example is CGCGTT, found in the promoters of genes which rapidly respond to wounding (as well as other stresses) within an hour after treatment (Walley et al., 2007). Another example is the G-box, CACGTG, which is bound by Myc TFs in response to wounding and JA treatment after 1 h (Fernández-Calvo et al., 2011). Other elements implicated in early wound response include W-box (TTGACC), GCC box (AGCCGCC or GCCGCC), jasmonate and elicitor-responsive expression element (AGACCGCC) and drought-response element (TACCGACAT; Rushton et al., 2002; Godoy et al., 2011). Most wound regulatory element studies focus on response to wounding after 1 h or validation of the derivatives of the G-box element, with only a few studies focusing on later time-points that are not G-box related (He and Gan, 2001). Studies that have compared early and late time points have focused on changes in gene expression over time (Ikeuchi et al., 2017), not on how regulatory elements change or have found only regulatory elements related to the G-box element across time points (Delessert et al., 2004). There are likely additional CREs that remain to be discovered at both early and late time points. Notably, the earlier studies have demonstrated the feasibility of establishing computational models capable of predicting plant spatial stress responses to, e.g. high salinity, using known and newly discovered CREs (Uygun et al., 2017). In doing so,

the computational model provides a means to globally assess the extent to which included CREs are sufficient to predict stress response and to pinpoint the most relevant CREs. Nonetheless, such a model is yet to be established for wound response. Thus, while several studies have found specific regulatory elements related to wounding, a global model of wound-response regulation across time has not emerged. In addition, wound response has been shown to have JA-dependent and JA-independent components (Schillmiller and Howe, 2005). However, existing studies on this area focus on comparing wound-inducible genes that are JA-dependent or JA-independent (e.g. Reymond et al., 2000), roles of regulators controlling JA-independent response (Stotz et al., 2013), signaling components of JA-dependent response involving COI1 (Bömer et al., 2018). There is not yet a study comparing the cis-regulatory program of JA-dependent and JA-independent response.

The goals of this study were to uncover the cis-regulatory code involved in regulating temporal responses to wounding stress, to see how wounding stress independent of the wound-induced hormone JA is regulated, and finally to understand how genes in certain specialized metabolism pathways are regulated. Here, we assessed the extent of divergence in gene expression among various time-points following wounding by correlating wounding data with other types of stress or hormone treatment using an existing modeling framework (Liu et al., 2015; Uygün et al., 2017). By using a time course data set, where transcriptional response was recorded over a 24-h period (Kilian et al., 2007), we captured differences in differential gene expression among varying time-points and the global regulatory pattern required to regulate these transcriptional responses. While this data set has been used before to identify regulatory elements (Ma and Bohnert, 2007), the CREs identified were limited to only known motifs from the PLACE database (Higo et al., 1999), and the focus was on general stress response. In this study, we expand on earlier studies by including CREs derived from model predictions, applying a system-wide modeling approach, and focusing on responses to specific stress i.e., wounding. Because most regulatory elements occur 1,000-bp upstream of the transcriptional start sites in the promoter region of the gene in Arabidopsis (Weirauch et al., 2014; Yu et al., 2016), we focused on this region to identify putative CREs (pCREs). In addition, by clustering wound-responsive genes into groups based on whether they respond to JA, we were able to single-out differences between JA and non-JA regulatory mechanisms regarding wounding. Furthermore, we identified important regulatory elements for the wound-responsive genes in the pathway glucosinolate biosynthesis from tryptophan, which is induced by wounding. Finally, by using machine learning modeling, we were able to identify the most important regulatory elements for each time-point and experimentally validate one known and three previously unknown elements regulating wounding response.

Results and discussion

Transcriptional response to wounding across time-points

To understand how transcriptional response to wounding varies across time-points, we used the wounding treatment data from an existing expression data set, which contains seven abiotic stress treatments and where the wounding treatment was applied to 18-day-old Arabidopsis seedlings (Kilian et al., 2007). Samples were harvested at multiple time-points ranging from 15 min to 24 h after wounding treatment. The sampling of control treatment was performed in parallel to exclude circadian effects (Kilian et al., 2007). We identified genes that were up- or downregulated at the different time-points (diagonal values; Figure 1A) and how frequently the same genes were differentially expressed in these different time-points (lower triangle; Figure 1A). The type of wound-responsive genes was named with two components: (1) time-point after wounding, (2) up- or downregulated after wounding treatment, e.g. 0.25hr_up, 1hr_down (Figure 1A). There was a cascading effect, where the majority of 0.25hr_up and 0.5hr_up genes overlap with each other (84% and 61%, respectively) and were still upregulated at 1 h (63% and 70%, respectively), but by 3 h $\leq 25\%$ of those genes were still upregulated (Figure 1A, also see Supplemental Data Set S1 for genes present in each wound-responsive cluster). Thus, different time-points after wounding have overlapping but distinct sets of genes, which are up- or downregulated, suggesting temporal variation in how wounding response is regulated.

To determine how response to wounding differs from response to other environmental conditions, we measured how similar the pattern of differential gene expression was between different wounding time-points and other abiotic stress, biotic stress, and hormone treatments (see “Materials and Methods”). The Pearson’s correlation coefficient (PCC) was used to measure the correlation of the log₂ fold change (log₂FC) values across genes between wounding and other stress/hormone treatments. The PCC values of genes were used to group treatment data sets with similar differential gene expression using hierarchical clustering (Figure 1B). We found that wound response correlated with both abiotic and biotic stress responses (Figure 1B). Early patterns of wounding differential gene expression (DGE; 15, 30 min or 1 h after wounding) were more highly correlated with those of early abiotic stress response compared with later wounding time-points (12 or 24 h after wounding). Gene expression patterns at 30 min and 1 h after wounding were also more similar to early responses (30 min to 3 h) under abiotic stresses, such as cold, UV-B, osmotic, and genotoxic stress (Figure 1B; Supplemental Data Set S2). Additionally, early DGE 15 min and 1 h after wounding were more similar to each other (PCC = 0.39) and to 30 min after wounding (PCC = 0.33 and 0.30, respectively) than to later wounding time-points (for PCC results, see Supplemental Data Set S2). Thus, transcriptional responses were more similar among some comparable time-points between treatments than

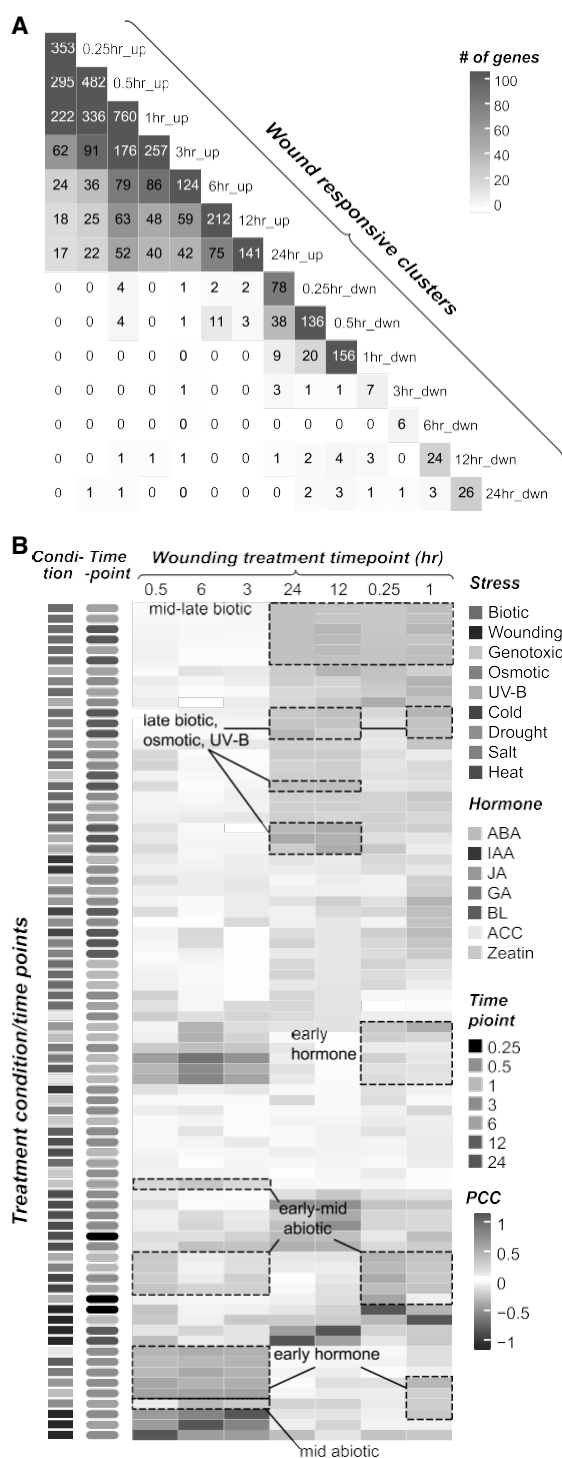


Figure 1 Gene expression correlation across stress and hormone data sets and the overlap of wound and JA differentially expressed genes. **A**, Heatmap showing the number of genes overlapping in each wound-response cluster. The order of rows and columns are the same, based on time-points and directions of differential regulation. Number of genes range from 0 (white) to 760 (red) and actual values are provided in the heatmap. **B**, Heatmap of PCC based on the log₂FC between treatment and control among different conditions (stress or hormone treatment) at different time-points. PCC values in heatmap range from 1 (red) to -1 (blue). The rows and columns were ordered based on hierarchical clustering. The stress and hormone treatments as well as treatment time-points are labeled by colors.

among largely differing time-points within a particular treatment. This indicates that temporal patterns can influence gene expression more than the type of abiotic stress, and that wounding can elicit a similar response to other types of abiotic stress. When observing wounding-response patterns in relation to biotic stress, 15-min, 1-, 12-, and 24-h time-point wounding responses all correlated with different types of biotic stress (Figure 1B; Supplemental Data Set S2). Thus, early wounding response is correlated with both abiotic and biotic stress. However, at later time-points (12 and 24 h), wound responses are more highly correlated with late biotic stresses than with any other stress (PCC for biotic stress *P. infestans* at 12 h was 0.48 and 0.38 for 12 and 24 h). On the other hand, wounding DGE responses at 3 and 6 h after wounding do not correlate with biotic stress response (PCC range -0.09 to 0.07; Supplemental Data Set S2). Our findings confirm that the initial wounding response is akin to general stress response as previously suggested in Walley et al. (2007).

In terms of the relationships between wounding and hormonal responses, DGE 15 min after wounding was not similar to DGE 30 min after hormone treatment (PCC range -0.07 to 0.14 for all treatments). By 30 min after wounding, however, DGE was similar to DGE 30 min after treatment with ABA, amino-cyclopropane carboxylate (ethylene precursor, ACC), brassinosteroid (BL), gibberellic acid (GA), and JA, PCCs ranging from 0.37 to 0.52 (Figure 1B; Supplemental Data Set S2), indicating that initial response to wounding triggers the production of multiple hormones. The DGE responses at 3 and 6 h after wounding were even more similar to the DGE response after 30-min treatment of ABA, ACC, BL, GA, and JA (PCC range, 0.39–0.54), than were most other wounding time-points (PCC range, -0.04 to 0.21; Figure 1B; Supplemental Data Set S2). Finally, 12 and 24 h after wounding, transcriptomic responses showed little correlation with DGE responses after 30 min of hormone treatment (PCC range, -0.15 to 0.26; Supplemental Data Set S2). Overall, the high correlations of DGE patterns in early and 3- to 6-h time-points after wounding to early hormone treatment suggests wounding activates a hormonal response, recruiting hormone-responsive genes among other genes.

Modeling temporal wound response using machine learning

The temporal differences in transcriptional response to wounding described above suggest that the regulation of wounding-response changes over time, with regulatory control being more similar within early and mid-range time-points (0.25, 0.5, 1, 3, and 6 h), and within late time-points (12 and 24 h) compared to between these time-points (Supplemental Data Sets S1 and S2). Thus, wound-responsive genes were divided into 12 clusters depending on each time-point and the directions of response, for example 1hr_up refers to being upregulated at 1 h while 3hr_down refers to being downregulated at 3 h (Figure 1; see

“Materials and Methods”). To compare what regulatory mechanisms were important across different time-points/response directions, we estimated the regulatory code of transcriptional response to wounding for each cluster using machine learning approaches. Here, the regulatory code for a cluster was defined as a machine learning model that could classify a gene as being differentially regulated or non-differentially regulated in a cluster based on likely regulatory sequences. Note that the regulatory code of downregulation 3 and 6 h after wounding were not modeled because too few genes (≤ 10) were in these clusters.

First, we tested how well known regulatory sequences were able to model wounding response. We collected 52 known cis-regulatory elements (referred to as CREs) associated with JA, wounding, or insect responses identified previously using experimental or computational approaches (see “Materials and Methods”; Supplemental Data Set S3). We mapped each putative regulatory sequence to the putative promoter regions (see “Materials and Methods”) of each gene in a cluster, as well as to genes in a “null” cluster, consisting of genes that are not significantly upregulated or downregulated under any stress or hormone treatment. Two algorithms, random forest (RF) and support vector machine (SVM) were used to build models for each wounding-response cluster using cross-validation (see “Materials and Methods”). In all sections, RF results were reported unless noted otherwise. To measure model performance, F-measure was used which jointly considers precision and recall (see “Materials and Methods”). Using known CREs, the F-measures for models built for each wound-response cluster ranged from 0.67 to 0.71 (median = 0.68), scores that show our models performed better than random guessing (F-measure = 0.5) but were not perfect predictors (F-measure = 1; for RF models: Figure 2A, for SVM models: Supplemental Figure S1A and Supplemental Data Set S4). Note that the cluster 12hr_down was not analyzed because no known regulatory elements were defined as present in the promoters of the genes in this cluster.

Next, we incorporated additional regulatory information to see if our model could be further improved. We included in vitro DNA binding data of 510 TFs in Arabidopsis generated with DNA affinity purification sequencing (DAP-seq; O’Malley et al., 2016) and information about DNase I hypersensitive sites (DHSs) in Arabidopsis sampled at different developmental stages including seedling (leaf samples) and 2-week-old plants (flower buds; Zhang et al., 2012). Each DAP-seq and DHSs feature was considered present if its peak coordinates overlapped with the promoter region of a gene. Models trained using both known sequence and DAP-seq and DHSs features generally performed slightly better than models using only known CRE, with the F-measure ranging from 0.66 to 0.74 (median = 0.69; Figure 2; Supplemental Data Set S4). Models for genes upregulated in early wounding response (0.25, 0.5, and 1 h) benefited the most from the addition of these two data sets, with an

increase of 0.03, 0.03, and 0.02 in F-measure, respectively. This may be because early wound-response clusters have larger gene numbers than clusters for later time-points. Having a larger gene set may allow for a higher degree of overlap with DAP-seq or DHSs features. Thus, more known information in the form of the DAP-seq data may improve the performance of early time-point clusters more than later time-points. Overall, while known sequence-based information together with DAP-seq and DHSs information is predictive of differential gene expression in response to wounding across time-points, the models still have substantial room for improvement.

Determining the relative importance of known motifs and additional regulatory information for predicting temporal wound response

To understand what known elements, TFs (based on DAP-seq), and DHSs are particularly important for predicting responses at different times after wounding, we determined the importance of each feature in each model (see “Materials and Methods”; Supplemental Data Set S5). In Figure 2B, the top 10 features for upregulated time-point clusters are shown. For early wound response (genes upregulated 0.25, 0.5, and 1 h after wounding), the most important known CREs were CGCGTT (first ranked), a known regulatory element for rapid wound response (RWR; Walley et al., 2007) and CACGTG (second ranked) that is bound by TFs in the bHLH family in response to wounding and JA treatment (Fernández-Calvo et al., 2011). Genes with the RWR elements are known to respond quickly to wounding and have a variety of functions in the downstream responses, including chromatin remodeling, signal transduction, and mRNA processing (Walley et al., 2007). TFs that respond to wounding stress such as MYC2, MYC3, and MYC4 bind the CACGTG motif and respond to both JA and wounding, and induce other JA-responsive genes, ultimately triggering defense response to herbivory (Fernández-Calvo et al., 2011). In addition to the important contribution to the regulatory code for genes upregulated 0.25–1 h post wounding, CACGTG was still important (ranked 1 or 2) among genes upregulated 3, 6, and 12 h after wounding, while the RWR element was no longer the most important contributor. By 24 h after wounding, the CACGTG element was no longer highly ranked.

DAP-seq binding sites were less important in predicting wound response than the known CREs or DHSs (Figure 2B; Supplemental Data Set S5). A few DAP-seq binding sites ranked among the top 10 most important features, including the calmodulin binding transcription activator (CAMTA) TFs that bind to AAGCGCGTG and were ranked third most important for genes upregulated 0.25 or 0.5 h after wounding but dropped to 11th at 1 h after wounding, and were even lower in later time-points. Consistent with earlier findings, CAMTA TFs are general stress-response factors triggered early during multiple stresses, including wounding

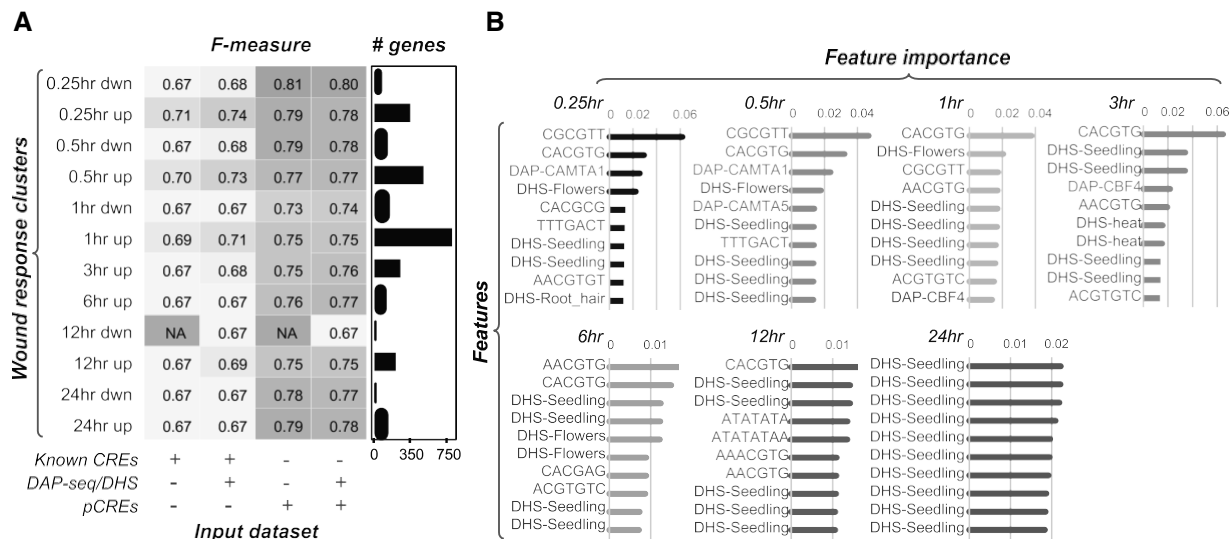


Figure 2 Performance of wound-response cluster prediction models. A, Each row indicates models for a specific wound-response cluster using different input features (columns). The models were generated with RF. Known CREs refers to those reported in the literature (Supplemental Data Set S3). The F-measure ranges from 0.5 (white) to 1 (red). The bar chart next to the heat map represents the numbers of genes in clusters. B, The top 10 most important features in models built using known CREs, DAP-seq sites, and DHSs for upregulated time-point clusters. The bars are colored in the same way as Figure 1B.

(Benn et al., 2014). The CBF TF in the AP2/EREBP family that binds to GGCGGCGGCGG ranked 10th in the 1 h/ upregulation model and 4th at 3 h after wounding. Interestingly, it has been reported that CAMTA TFs regulate CBFs (Doherty et al., 2009). Thus, the ranked importance of these TFs at each time-point is consistent with their regulatory interactions. Nonetheless, all DAP-seq sites became less important for predicting genes upregulated 6, 12, and 24 h after wounding. Because known CRE sites in response to wounding were ranked in these clusters but DAP-seq sites were not, this may reflect the fact that DAP-seq identifies TF binding sites (TFBSs) *in vitro*, regardless of whether the sites are accessible *in vivo* or not. Together with the mediocre model performance (Figure 2A), these findings show temporal differences in wounding regulatory codes, but also that known CREs and DAP-seq data do not fully capture how wounding response is regulated, especially at later time-points.

In addition to known CREs and DAP-seq sites, open chromatin sites (DHS) were important for predicting expression regulation at all time-points after wounding (top-ranked DHS sites for each cluster ranged from ranks 1–4), particularly at later time-points (Figure 2B; Supplemental Data Set S5). For example, at 24 h after wounding, the top 12 most important features were all DHS-related. We hypothesize two potential explanations for this finding. First, at later time-points, the functional diversity of expressed genes has increased so that their transcriptional regulatory mechanisms have become more complicated (due to both wounding and repair mechanisms) and thus no single CRE or DAP-seq feature can be found with high importance. The second possibility is that the known CRE or DAP-seq features

important for later time-points are not present in our data set. Although important, DHS sites do not provide additional information to improve the F-measures of our models especially at later time-points. Thus, we hypothesized that regulatory sequences not yet identified could be important regulators of wound response, especially for later wound response.

Finding important temporal putative *cis*-regulatory elements for wound response

To test our hypothesis that there were unknown regulatory sequences controlling wounding response, we identified pCREs with a *k*-mer finding approach (Liu et al., 2018), where all possible 6–30-mer sequences were tested for enrichment in the putative promoters of genes for each wound-response cluster (see “Materials and Methods”). Based on this criterion, 42–1,081 pCREs were identified as enriched in genes from each wound-response cluster, with the exception of the 12hr_down cluster, which had no enriched pCREs (for enrichment statistics of pCREs; see Supplemental Data Set S6). For each wound-response cluster, the pCREs were used to build five replicate wound-response prediction models and the reported model performance (F-measure) and feature importance were based on averages of the five models. We found that models built with pCREs alone (F-measure range = 0.73–0.81; Figure 2A; Supplemental Data Set S4) perform better than models built with known CREs, DAP-seq and DHSs for all clusters (F-measure range = 0.66–0.74; Figure 2A; Supplemental Data Set S4). Because the number of pCREs exceeds the number of known CREs, we modeled the 1hr_up wounding cluster using only the top 52 pCREs and compared model

performance to the model using the 52 known CREs. We found that the top 52 pCRE-based model performs slightly better (F-measure of 0.72) compared to the known 52 CRE-based model (F-measure of 0.69). Interestingly, models that were built by combining pCREs with DAP-seq, and DHS data (F-measure = 0.67–0.80; Figure 2A; Supplemental Data Set S4) did not necessarily perform better than models built using only pCREs, with the exception of the 12hr_up time-point, perhaps reflecting the increasingly more important roles of regulation beyond the cis-regulatory level. We also note that the 12-h and 24-h time-points, as well as downregulated gene clusters, have smaller gene numbers overall, thus while they have high F-measures, this may make these models less generalizable than models with higher gene numbers. In addition to building binary models, we also built a regression model for the 1-h time-point to see if the level of expression could be predicted using regulatory features. The regression models, however, were not predictive (for SVM = PCC –0.07, RF = PCC –0.44; Supplemental Data Set S4), revealing the challenges in predicting expression level but also the benefits of clustering genes based on their expression to make better binary predictions. Overall, these findings indicate that these pCREs contributed information beyond what was available from known DAP-seq and DHSs data in the regulation of wound response at different time-points.

To understand why the models improve by using pCREs, and what influence pCREs have across wounding time-points relative to known information and open chromatin sites, we looked at the average importance rank (normalized importance score, scaled between 0 and 1, see “Materials and Methods”) of all features in the models (including features related to pCREs, DAP-seq sites, and/or DHSs) across the post-wounding time course (Figure 3). We found that DHSs tend to be the most important features for most time-points (Figure 3, A–J) apart from late downregulated time-points (Figure 3, K and L). However, in each of these clusters we found some pCREs to be more important than DHS features. For example, at 1hr_up 169 out of the top 200 features were pCREs. Finally, DAP-seq sites were less important than DHSs and pCREs except at 12-h and 24-h time-points for downregulated genes. Although the DHS data used was not generated under wounding stress (Zhang et al., 2012), it is surprisingly useful and we cannot rule out the possibility that the plants were actually wounded during sample preparation. To completely capture the importance of chromatin accessibility in wound response, we will need DHS data generated with wounded plant samples instead of using DHSs data generated in a different context. We should also note that DAP-seq sites are always of the least importance. With our findings that adding DHS/DAP-seq information does not improve our models (Figure 2A) and the fact that pCREs are also important at every time-point, this indicates that the identified pCREs may better uncover the regulatory code complexity underlying wound-response

regulation than, particularly DAP-seq sites that are available for a subset of TFs in Arabidopsis.

Correlation to TF families and *cis*-regulatory differences across time

Figure 4 shows the importance rank across all time-points for the top 10 most important pCREs for each wounding model. Like how similar sets of genes were differentially expressed at nearby time-points (i.e. the cascading effect), we found more important pCREs were shared between closer time-points (Figure 4). Because the pCREs were discovered from genes with similar wound-response patterns, this cascading effect on the shared number of important pCREs was expected. However, some pCREs were uniquely important for a narrow time frame (for the importance rank of pCRE and PCC of pCREs with known TF binding motifs (TFBMs), see Supplemental Data Set S7; for raw importance scores, see Supplemental Data Set S8). Next, we determined which pCRE was similar to a known TFBM and which was likely a previously unknown regulatory element. We first calculated the sequence similarity between each pCRE and each known binding motif. For this we used DAP-seq sites, which are generated *in vitro*, as well as CIS-BP sites (Weirauch et al., 2014), which are TFBSs found *in vivo* using chromatin immunoprecipitation sequencing.

For early time-points after wounding (i.e. 0.25, 0.5, and 1 h), many of the top important pCREs were shared and resembled TFBSs in the CG-1/CAMTA, bZIP/BZR, FAR1, LOB, and bHLH TF families (right two panels; Figure 4; Supplemental Data Set S7). The finding that different binding sites resemble multiple TF families is consistent with the notion that a variety of signals, and thus TFs, are induced by wounding (Howe, 2004; Zhang et al., 2016). However, not all pCREs were highly correlated with known TF binding families as 26 correlations between DAP-seq sites and pCREs, and 77 correlations between CIS-BP sites and pCREs shown in Figure 4 had PCCs ≤ 0.75 , indicative of substantial differences between pCREs and known TFBSs. Focusing on the top 3 most important pCREs for each time-point (Figure 5, based on average rank—see “Materials and Methods”), the pCREs CCGCGT and CACGTG were most similar to the binding motifs of CG-1/CAMTA and MYC2 bHLH TFs, respectively. Thus, the timing when they were important for predicting wounding response was consistent with the timing when known CREs for CAMTA and MYC2 TFs were important for the models built using only known CREs, DAP-seq, and DHSs (Figure 2B). The CACGTG element, which is important in binding TFs for JA response (Fernández-Calvo et al., 2011), remained important at both 3 and 6 h after wounding (ranked 10 and 5, respectively), indicating JA responses have been activated. Other top 3 important early pCREs remained important across the wider range of time-points and were not known as wounding CREs. One example is ACACGT, a pCRE most similar to the binding motif for bZIP family TFs, which are activated by ABA (Yamamoto et al., 2011) and regulate responses to water deprivation

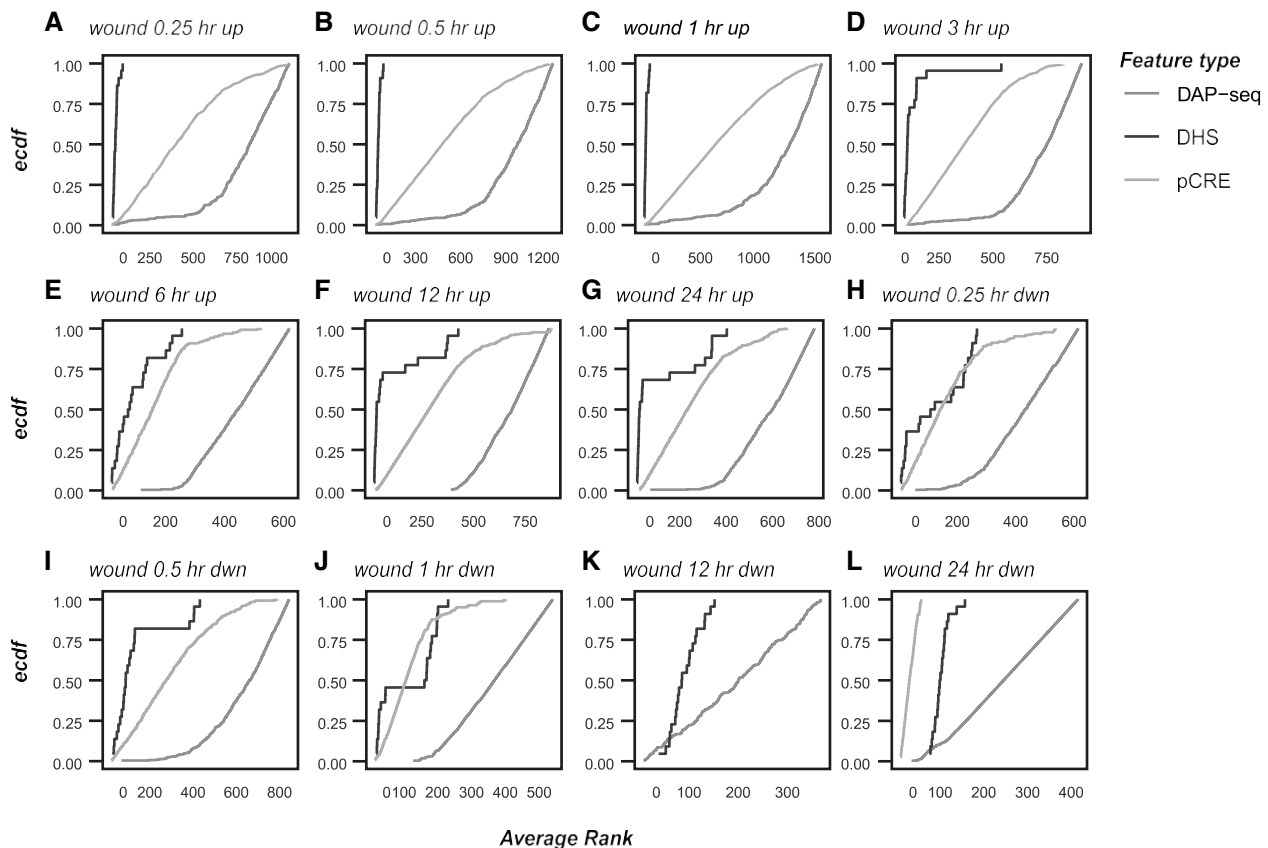


Figure 3 Scaled importance values for each wound-response model (rows) for all features used in the final model. For (A)–(L), density plots show normalized importance value on the *x*-axis and the density of the features on the *y*-axis. The importance value is scaled from 0 to 1 for each model. Only the top 200 features are shown, thus the *x*-axis varies depending on how many features in total there are for a given time-point. DAP-seq sites are in purple, DHSs are in yellow, and pCREs are in blue. The higher value correlates with higher importance of a given feature for a given model. (A–G) models for upregulated wounding-response clusters at (A) 0.25, (B) 0.5, (C) 1, (D) 3, (E) 6, (F) 12, and (G) 24 h. (H–L) models for downregulated wounding-response clusters at (H) 0.25, (I) 0.5, (J) 1, (K) 12, and (L) 24 h.

(Figure 5). This pCRE was enriched in the promoters of genes from all time-points and important (rank ≤ 11) for models of wounding response at all time-points except 24 h (Figure 4; Supplemental Data Set S7). Two pCREs (GTCGGC and GTCACA) that did not resemble known wounding CREs were uniquely important for models built for mid-range time-points (i.e. 3 and 6 h after wounding), as the 5th most important pCREs for the genes upregulated at 3 h and 6th most important at 6 h, respectively (Figure 5). These elements were most similar to binding motifs of B3 and Homeodomain family TFs, respectively. Given these TF families are involved in development, response to auxin, and secondary wall biogenesis, this indicates that by 3–6 h after wounding, the damage is likely being repaired. At the last two time-points (12 and 24 h after wounding), ATATTAT, which was most similar to binding motifs of TFs in the ARID family, was ranked 24th and 14th respectively (Figure 5; Supplemental Data Set S7). The ARID family is involved in regulating glucosinolate metabolism, indicating that specialized metabolism pathways are turned on or augmented 12 h after wounding and are still important after 24 h. Another two

important pCREs at the latest time-points, ATAATAA and AAAATGT, resemble elements that were bound by TFs from Homeodomain and GRF (Growth-regulating factors) families, respectively (Figure 5; Supplemental Data Set S7), which regulate development, which may be important in repairing the wound (van der Graaff et al., 2009; Omidbakhshfard et al., 2015; Wu et al., 2019).

In summary, we found that pCREs important for our models contain some known wounding CREs, but mostly regulatory sequences that are not known to be involved in wounding response. Additionally, while similar to TFBSs, most pCREs are not identical and contain slight changes in key positions, which may affect binding specificity. pCREs important for wound-response models at early time-points (0.25–0.5 after wounding) tend to be associated with multiple stress and hormone responses, while pCREs important for models 1–h after wounding tend to be associated with TFs involved in JA and ABA signaling. Finally, 3–24 h after wounding the important pCREs tend to be associated with TFs involved in growth and pCREs important for very late responses (12–24 h after wounding) are associated with some TFs related to metabolic defense. Our models of the

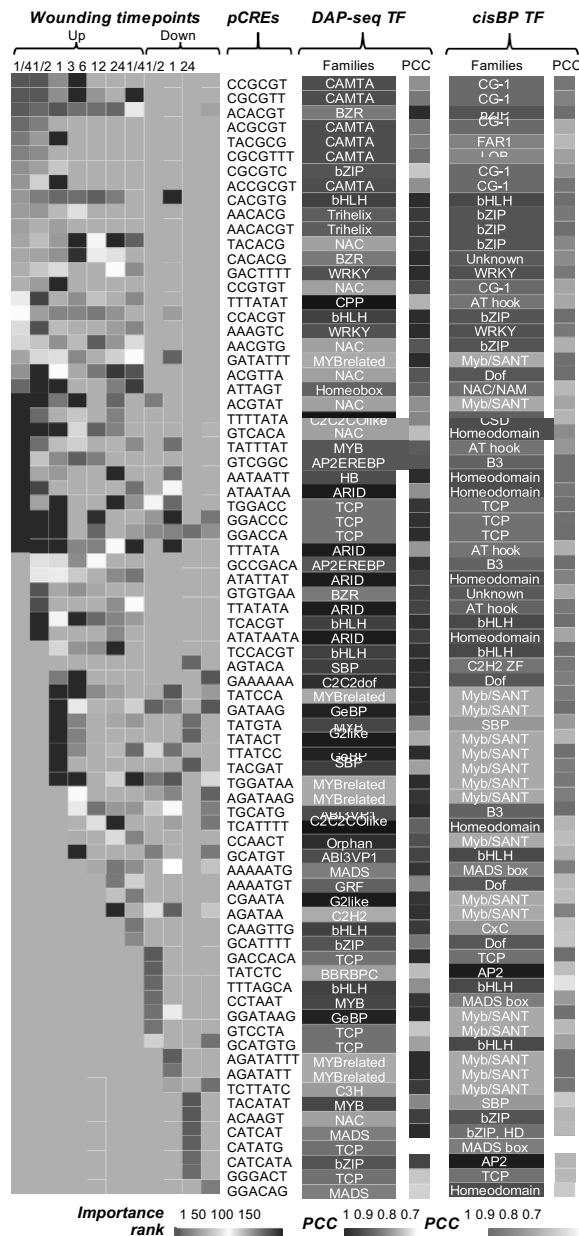


Figure 4 Average importance rank for the top 10 pCREs for each wound-response model and their association to a TF family. Wound-response models are the columns while pCREs are the rows. The top 10 pCREs for each model are shown, and how those pCREs overlap with other models in terms of importance rank. The average importance rank shown is the rank of average importance of a feature across five duplicate models ran for the same time-point. Highest rank (1) is red and ranks 150 or lower are blue. Gray color indicates that the pCRE is not present at that wound-response time-point. Association between pCREs and TF families was based on the similarity (measured using PCC) between sequences of pCREs and the previously reported binding sites (identified by DAP-seq or cis-BP) of TF families. The TF family with the maximum PCC to a pCRE was associated with the pCRE in question. PCC is shown for both DAP-seq (degrees of blue color) and cis-BP (degrees of green color) sites.

cis-regulatory code in response to wounding demonstrate how sets of pCREs, which are likely bound by a variety of TFs, are important at different response times after

wounding and could work to regulate a dynamic response to wounding over time.

Experimental validation of important CREs in early wound response

We validated our findings using the CRISPR/Cas9 system in *Arabidopsis* to evaluate the biological significance of two of the important pCREs (CCGCGT and CACGTG) based on our model and prior studies. CCGCGT is the top (most important) pCRE found for models of 0.25 (rank = 1), 0.5 (rank = 1), and 1 h (rank = 3) after wounding (Supplemental Data Set S7). CACGTG, a known CRE involved in wounding response (Figuerola and Browse, 2012), is ranked 30, 17, and 8 at 0.25, 0.5, and 1 h after wounding (Supplemental Data Set S7). CCGCGT is a variation of the CGCG box, that has been previously characterized as responsive to wounding signals, as well as other hormone (ABA) and oxidative signals, by binding the TF CAMTA5 (CALMODULIN-BINDING TRANSCRIPTION ACTIVATOR 2, AT4G16150) to the ethylene-responsive gene *EIN3* (ETHYLENE-INSENSITIVE3, AT3G20770; Yang and Poovaiah, 2002). The CAMTA-related TF family and its binding to the CGCG box has also been shown to respond to cold treatment and may impart freezing tolerance in *Arabidopsis* (Doherty et al., 2009). However, it is unknown how this motif affects other wound-responsive genes and has not been assessed in vivo via CRISPR-Cas9. Thus, we chose CRISPR/Cas9 target promoters that contain the CCGCGT motif and sorted out the candidates by the following criteria. First, the expression level of the gene, which has the target motif on its promoter region, was relatively high in fold change at the early time-points in response to wounding stress. Second, the pCRE site is near the PAM sequence such that the site renders susceptible to the CRISPR/Cas9 mutation (Osakabe et al., 2016). We finally selected genes *JAZ2* (*JASMONATE-ZIM-DOMAIN PROTEIN 2*, AT1G74950) and *GER5* (*GEM-RELATED 5*, AT5G13200). *GER5*, although not known to be involved in wounding response, was highly expressed 0.5, 1, and 3 h after wounding and contained the CCGCGT motif in the promoter region. *JAZ2* is a well-known JA-responsive gene (Fernández-Calvo et al., 2011) and the promoter region contained the G-box motif (CACGTG), which can be utilized as a positive control in our mutation assay (Figuerola and Browse, 2012), as well as the CCGCGT motif.

Next, we made the CRISPR/Cas9 construct that targets the pCREs in *JAZ2* and *GER5* promoters and transformed it into *Arabidopsis* with the Col-0 background. From antibiotic resistance T₁ plants, we found a homozygous mutant called *jaz2-4 ger5-3*. The *jaz2-4 ger5-3* mutant had one base pair insertion in the CCGCGT motif on both *JAZ2* and *GER5* promoters, where T insertion in the *JAZ2* promoter led to no significant nucleotide change from CCGCGT to CCGCGTT, while G insertion in the *GER5* promoter caused a base alteration from CCGCGT to CCGCGGT (Figure 6). Further, we generated a homozygous mutant called *jaz2-5* that harbored a mutation within the G-box motif of *JAZ2* promoter, in

which the CACGTG motif was mutated to CACGTTG (Figure 6). To determine the effect of the motif mutations on their downstream gene expression upon wound treatment, we harvested the seedlings of *jaz2-4 ger5-3* and *jaz2-5* mutants, as well as Col-0 controls, 1 h after wounding. The transcript abundances of both mutants and Col-0 were analyzed by reverse transcription quantitative polymerase chain reaction (RT-qPCR).

In the *jaz2-4 ger5-3* mutant the expression of the *JAZ2* gene was upregulated after wounding, exhibiting the same phenotype as the Col-0 control (Figure 6). Thus, as expected, regulation of *JAZ2* was not altered in *jaz2-4 ger5-3* by wounding, because the CRISPR–Cas9 mutation had resulted in a synonymous change. Although this in itself does not show that the G-box affects wounding, it is what was expected from the CRISPR result since CACGTT is also considered to be a G-box, and this was previously shown to be important for regulating JA/wound expression (Figueroa and Browse, 2012). Interestingly, the expression level of *GER5* was not changed in *jaz2-4 ger5-3* upon wound treatment, while the expression of *GER5* was significantly upregulated in the Col-0 control (fold increase = 5.52). This indicates that CCGCGT, a derivative of the stress-responsive motif CGCG-box, enables the *GER5* gene to respond to early wounding and is disabled by the G insertion.

The *JAZ2* expression was upregulated in response to the wounding treatment in both Col-0 and the *jaz2-5* mutant. Additionally, the *GER5* transcript level increased after wounding in the *jaz2-5* mutant (Supplemental Figure S2). In the case of *JAZ2*, the G-Box CACGTG was changed to CACGTT, which is a G-Box variant (Dombrecht et al., 2007). Thus, while significant, the change did not substantially alter the *JAZ2* response (fold increase = 1.04) compared with the *jaz2-4ger5-3* mutant (fold increase = −0.08). The *GER5* expression was also not markedly different in the *jaz2-5* mutant relative to the Col-0 in response to wounding (fold increase = −1.73, $P = 8.07\text{E-}05$). Taken together, these results indicate that the CCGCGT CRE is responsible for the wounding response of *GER5*.

Experimental validation of important unknown pCREs at later time-points

In addition to CREs important for early wounding response at or before 1 h, we validated important pCREs at 3 and 6 h time-points (see Supplemental Data Set S7) using amplified luminescent proximity homogeneous assay (ALPHA) in vitro DNA-binding experiments complemented by mutations in protoplasts coupled with a reporter gene to evaluate in vivo DNA binding (see “Materials and Methods”). For genes upregulated after 3 h of wounding, the most important pCRE to be GTCGGC, a site most similar to sites that bind the AP2EREBP TF family, and the next most important pCRE to be ACACGT, similar to BZR TFBSs. For genes upregulated 6 h after wounding, the most important pCRE is AACGTG, a derivative of the G-box motif (CACGTG) that also binds Myc TFs (Fernández-Calvo et al., 2011). Thus, we

decided to test the next most important unknown pCRE, GTCACA, which is most similar to NAC TFBSs. The next most important pCRE for the 6-h time-point is ACACGT, which was also important at three hours. In the end, we targeted three additional pCREs: GTCGGC, ACACGT, and GTCACA from the promoters of three genes AT5G07010, AT2G02990, and AT5G13220, respectively, for further testing.

For each pCRE, we first identified candidate TF of which the binding site is the most similar to the CRE sequence (see “Materials and Methods” and Supplemental Data Set S7). For the ALPHA assay, recombinant proteins for TF candidates were purified (Supplemental Figure S3A) and tested against their predicted motif for binding. For GTCACA and GTCGGC, more than one of the TF candidates were tested (Supplemental Data Set S9) where a subset did not bind (Supplemental Figure S3, B–D). For all three pCREs, the WT probes produced strong Alpha signals indicating protein binding to the DNA at all tested TF protein concentrations (Figure 7). In contrast, the probes containing the mutated pCRE sequences either did not produce signals or at a significantly lower level compared to WT probes (Figure 7). These findings indicate that the WT pCRE sequence is important for binding the candidate TFs.

Next, we checked in vivo binding using a protoplast assay for GTCGGC and GTCACA important for 3-h and 6-h post-wounding, respectively. For each gene containing either the GTCGGC or GTCACA pCRE in their promoters, we first measured mRNA accumulation levels in unwounded and wounded Arabidopsis plants and in Arabidopsis protoplasts (Supplemental Figure S4). We found expression for each gene increased after wounding compared to unwounded plants. In addition, we found expression in protoplasts for each gene exceeded the amount of expression in unwounded plants, and sometimes that of wounded plants. Therefore, transcript levels increased due to wounding was similar to the expression increase seen in protoplasts, relative to unwounded plants. We hypothesize that because of the similarity in gene expression there may be similarities in how the genes of wound response and protoplasts are regulated. Next, we assembled each pCRE site or its mutated version as a tetramer in a head-to-tail orientation upstream of a luciferase reporter gene (harboring a minimal CaMV 35S promoter, see “Materials and Methods”; Figure 8A) that was co-transfected into the Arabidopsis protoplasts. For both pCREs, we found significantly higher luminescence levels in protoplasts with WT sequences than with mutated ones (Figure 8B). This indicates that each motif is sufficient to induce expression in Arabidopsis protoplasts while the mutated motifs are not. Thus, the predicted GTCGGC and GTCACA motifs regulate reporter gene expression *in vivo*, providing evidence for their functionality in wound-response regulation, but stress that this does not experimentally show these pCREs to be directly involved in wound response.

Cluster	pCREs	Avg. rank	Best TFBM match		TF family	TF	GO categories
			Forward	Reverse complement			
0.25 hr-up	CCGCGT	1	PCC=0.97 		CG-1	AT2G22300	cellular response to cold, defense response to bacterium, defense response to fungus,
	CGCGTT	2	PCC=0.93 				
	ACACGT	3	PCC=1 		bZIP	AT3G19290, AT4G34000, AT4G18890	abscisic acid-activated signaling pathway, response to water deprivation, positive regulation of chlorophyll catabolic process
0.5 hr-up	CCGCGT	1	PCC=0.97 		CG-1	AT2G22300	cellular response to cold, defense response to bacterium, defense response to fungus,
	CGCGTT	2	PCC=0.93 				
	ACACGT	8	PCC=1 		bZIP	AT3G19290, AT4G34000, AT4G18890	abscisic acid-activated signaling pathway, response to water deprivation, positive regulation of chlorophyll catabolic process
1 hr-up	ACACGT	1	PCC=1 		bZIP	AT3G19290, AT4G34000, AT4G18890	abscisic acid-activated signaling pathway, response to water deprivation, positive regulation of chlorophyll catabolic process
	CACGTG	8	PCC=1 		bHLH	AT5G38860, AT1G32640	regulation of transcription, DNA-templated, transcription, response to JA, ABA signaling, defense response against insect, response to wounding
	CCACGT	21	PCC=1 		bZIP	AT1G49720, AT3G19290	abscisic acid-activated signaling pathway, positive regulation of transcription, DNA-templated
3 hr-up	GTCGGC	5	PCC=1 		B3	AT1G19220, AT1G19850, AT1G30330, AT5G20730, AT5G37020, AT5G60450	auxin-activated signaling pathway, lateral root development, leaf development, meristem development, response to auxin
	CACGTG	10	PCC=1 		bHLH	AT5G38860, AT1G32640	regulation of transcription, DNA-templated, transcription, response to JA, ABA signaling, defense response against insect, response to wounding
	ACACGT	11	PCC=1 		bZIP	AT3G19290, AT4G34000, AT4G18890	abscisic acid-activated signaling pathway, response to water deprivation, positive regulation of chlorophyll catabolic process
6 hr-up	AACGTG	4	PCC=0.92 		bZIP	AT3G04060	positive regulation of chlorophyll catabolic process, positive regulation of leaf senescence, regulation of transcription
	CACGTG	5	PCC=1 		bHLH	AT5G38860, AT1G32640	regulation of transcription, DNA-templated, transcription, response to JA, ABA signaling, defense response against insect, response to wounding
	GTCACA	6	PCC=0.9 		Homeo-domain	AT1G62990, AT4G32040, AT5G11060, AT5G25220	mucilage biosynthetic process, negative regulation of transcription, DNA-templated, regulation of secondary cell wall biogenesis, xylem development, response to ethylene, light, cytokine stimulus
12 hr-up	ACACGT	7	PCC=1 		bZIP	AT3G19290, AT4G34000, AT4G18890	abscisic acid-activated signaling pathway, response to water deprivation, positive regulation of chlorophyll catabolic process
	AAAAATG	15	PCC=0.98 		MADS	AT1G31140	fruit develop. & morphogenesis, integument develop., multicellular organism develop., neg. reg. of cell growth, plant ovule develop., pos. reg. of transcription by RNA pol. II, regulation from veg. to reproductive stage
	AAAATGT	16	PCC=0.85 		GRF	AT2G06200	leaf development, regulation of transcription, DNA-templated
24 hr-up	ATAATAA	12	PCC=0.97 		Homeo-domain	AT1G26960, AT1G69780, AT3G01220, AT4G40060, AT5G15150	lateral root formation, pos. reg. of transcription, DNA-templated, response to gibberellin, cotyledon morphogenesis, leaf morphogenesis, response to auxin, neg. reg. of cell growth, photoperiodism, flowering
	ATATTAT	14	PCC=0.95 		ARID	AT1G04880	glucosinolate metabolic process, pollen germination, pollen tube growth, regulation of transcription, DNA-templated
	CAAGTTG	15	PCC=0.97 		bHLH	bHLH80	N.A.

Figure 5 Motif logos for the top three pCREs for each upregulated wound-response cluster. Chart is divided by time-point (0.25–24 h after wounding). The first column is the top 3 ranked pCREs for each time-point. Note that ranking includes other features such as DHSs and DAP-seq sites, therefore the actual pCRE rank may be lower. The second column shows the average rank for a pCRE in the given model. The third and fourth columns show the best matched TFBM logos, with forward and reverse complement sequences, respectively. PCC values between pCREs and the TFBMs are indicated in the third column. Columns 5–7 are the TF that binds to a given binding site (column 6), the TF family the TF belongs to (column 5), and GO categories of the TF (column 7).

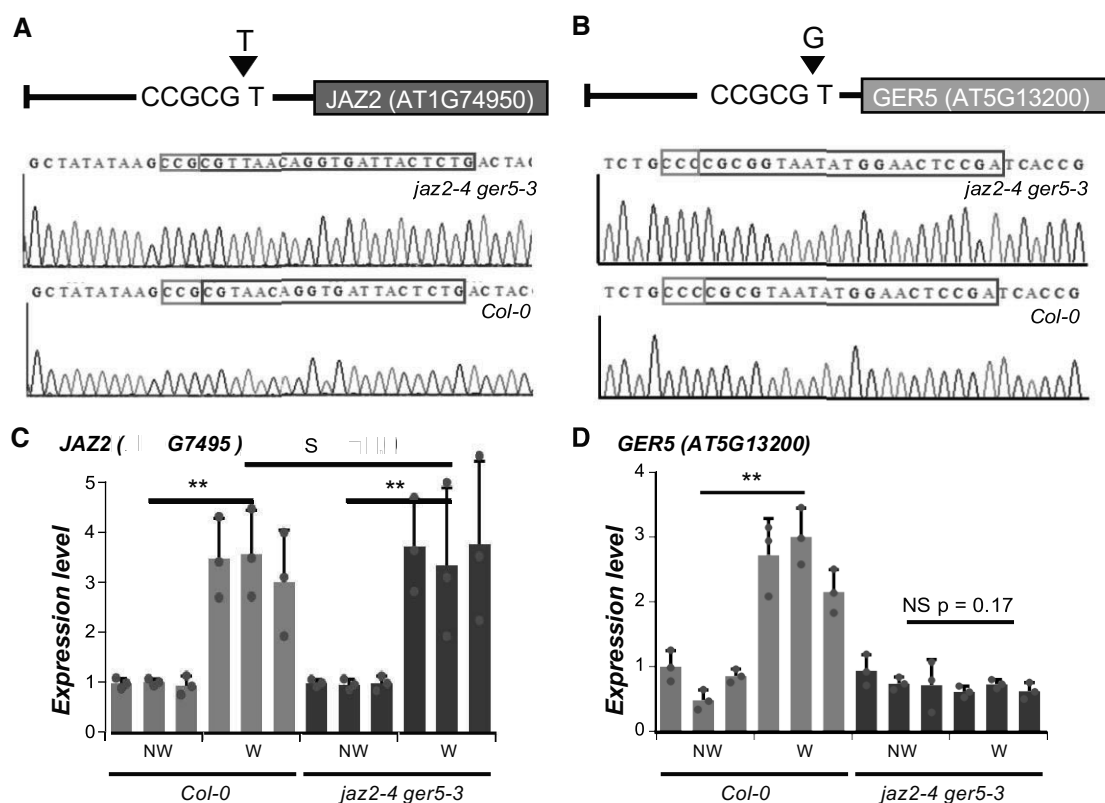


Figure 6 Mutation in predicted cis-regulatory motif abolished the induction of *GER5* by wounding. A, CRISPR/Cas9 mutation in the CCGCGT motif in the *JAZ2* promoter region (resulting in the *jaz2-4 ger5-3* line). B, CRISPR/Cas9-mediated mutation in the CCGCGT motif in the *GER5* promoter region in *jaz2-4 ger5-3*. Chromatogram represents the sequence of the *JAZ2* or *GER5* promoter region modified by CRISPR/Cas9 (upper chromatograms), and the corresponding region in *Col-0* (lower chromatograms). Blue and red boxes indicate PAM sequence and gRNA target regions, respectively. C and D, Wound responses of *JAZ2* (C) and *GER5* (D) expression in *Col-0* and in *jaz2-4 ger5-3*. Transcript abundances of *JAZ2* or *GER5* evaluated by RT-qPCR were normalized to *ACTIN2*. NW and W indicate no wound, collection after 1 h, and wound treatment after 1 h, respectively. Values for biological triplicates are shown using individual bars, while values for three technical repeats for each biological replicate were depicted with error bars. Significance levels of differences from the one-way analysis of variance were indicated with asterisks (Non-Significant [NS] $P \geq 0.05$, * $P \leq 0.05$, ** $P \leq 0.01$).

Modeling the regulatory code of JA-dependent and JA-independent gene response across wounding time-points

Having demonstrated pCREs important for predicting wound response, we next studied the regulatory differences between JA-dependent and JA-independent genes in the context of wound response. JA-independent wounding responses include those induced by RNase and nuclease activities that are triggered by wounding but not by the application of JA (LeBrasseur et al., 2002). Thus, to understand how JA-independent wound responses are regulated, we used the hormone treatment data (Goda et al., 2008) to identify wound-responsive genes that were also responsive to JA or not at 0.5, 1, and 3 h, for which data for both JA and wounding treatments are available. For these three time-points, 84%, 74%, and 72% of genes were upregulated after wounding but not after JA treatment, respectively (Supplemental Figure S5), consistent with the findings of a prominent JA-independent component in wounding response in other studies (León et al., 1998; LeBrasseur et al.,

2002). With this information, we divided the wound-response clusters from the 0.5-, 1-, and 3-h time-points into JA-dependent and JA-independent gene subclusters and generated model predicting wound response for each cluster using known CREs, DAP-seq sites, DHSs, and/or pCREs. Similar to our earlier results, pCRE-based models (F-measures: 0.73–0.87) outperformed both known CREs (0.67–0.74) and known CREs/DAP-seq/DHSs-based models (0.66–0.73; Figure 9; for SVM models: Supplemental Figure S1B and Supplemental Data Set S4). Thus, pCREs were able to better model the regulation of JA-dependent and JA-independent wounding response across time-points, than known TFBSs.

By comparing the importance of known CREs, DAP-seq sites, DHSs, and pCREs across models, we identified how JA-dependent and JA-independent responses differed in which known CREs and pCREs were important. At 30 min and 1 h after wounding, for example, CGCGTT, the RWR element, and CACGTG, the G-box recognized by many bHLH factors (Heim, 2003; Dombrecht et al., 2007), were the most important known elements for the JA-independent and JA-

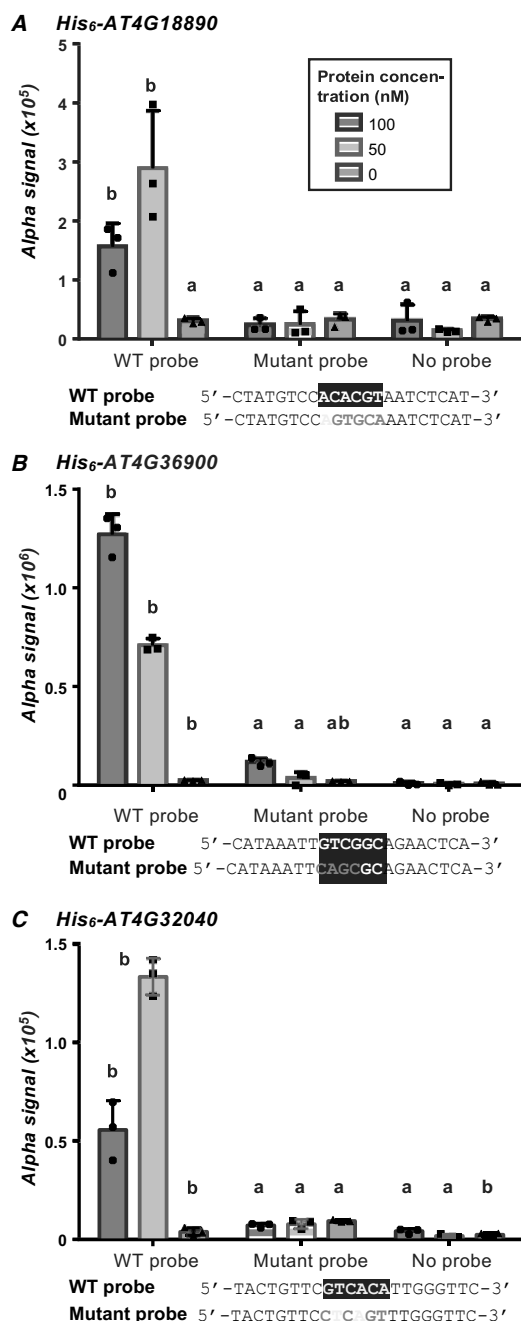


Figure 7 Binding of three TFs to identified cis-regulatory motifs. WT probes containing motifs ACACGT (A) from promoter AT2G02990, GTCGGC (B) from promoter AT5G07010, and GTCACA (C) from promoter AT5G13220, and their corresponding mutant probes were incubated with recombinant His6-AT4G18890, His6-AT4G32040, and His6-AT4G36900 proteins, respectively, and the DNA binding affinity was determined by ALPHA assay. Three different concentrations (0, 50, and 100 nM) of the proteins were examined with 10-nM probe in three technical replicates and two biological replicates. Similar trends of the results were obtained from the two biological repeats and one representative was shown with error bars indicating the standard deviation of technical replicates. The tested cis-regulatory motifs and the mutated sequences within the motifs are indicated shaded and underlined, respectively. The different letters indicate significant differences between groups evaluated by one-way analysis of variance followed by the Tukey's multiple comparison test at 5% significance level.

dependent models, respectively (see Supplemental Data Set S10 for pCREs, DHSs, and DAP-seq sites and their respective importance scores for JA-independent and JA-dependent models). Interestingly, the G-box element also ranks as the third most important feature in the JA-independent models. This could be because other TFs that are not involved in JA response (e.g. *Myc-LIKE* and *BIM3* TFs) can also bind to this element (O'Malley et al., 2016) or because the *Myc* element may be necessary to facilitate TF binding to a different regulatory element important for JA-independent response. For pCREs, with the exception of the G-box motif and the *bZIP* binding site (ACGTGT), there was little overlap in the ranking of important motifs between the JA-dependent and JA-independent models (Supplemental Figure S6). For example, AACGTG and CACGTTT were ranked from 1st to 7th across time-points in JA-dependent models but were not present or were ranked much lower (69th to 157th) for JA-independent models (Supplemental Figure S4 and Supplemental Data Set S8). In contrast, CCGCGT and GCCGAC were the most important pCREs 0.5 and 3 h after wounding in the JA-independent models but were not present or were ranked much lower (232nd importance) for JA-

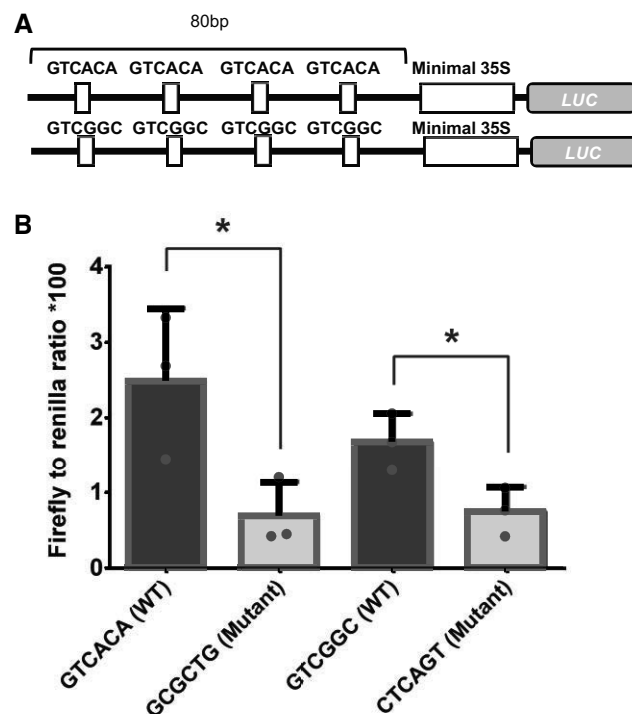


Figure 8 Mutation in the GTCACA and GTCGGC motifs attenuate expression of a reporter gene, when placed as tetramers upstream of a minimal 35S promoter. A, *Firefly luciferase* fused to the cis-regulatory motifs (WT or mutant–WT constructs are shown) were co-electroporated into Arabidopsis Col-0 protoplasts together with p35S:Renilla reporter, and luminescence levels were evaluated by dual bioluminescence assay. B, Luciferase activity was normalized by Renilla luciferase activity. Data represent mean \pm SD of three biological replicates of each WT and respective mutant construct, and an asterisk indicates $P \leq 0.05$ (Student's *t* test).

dependent models (Supplemental Data Set S10). Finally, we found that, of the top 10 most important features for each model, four to eight were DHSs for JA-independent models but none for JA-dependent models (Supplemental Data Set S10). Taken together, these findings highlight how JA-dependent and JA-independent responses were regulated by different sets of regulatory elements and were characterized by distinct chromatin accessibility patterns.

Modeling metabolic pathway regulation using wound stress data

We next assessed whether the regulatory sequences identified allow us to understand wounding response at the level of metabolic pathways. Here, we asked which specialized metabolism pathways were enriched in genes upregulated across the wounding time series (Supplemental Data Set S11). Depending on the time-point, 5–11 pathways were significantly enriched in wound-response genes. From 0.25 to 3 h after wounding, JA biosynthesis was the most highly enriched pathway (P -values range from $1.5\text{e-}3$ to $3.5\text{e-}7$; Supplemental Data Set S11). However, by 12 h it is no longer significantly enriched. Another example was the glucosinolate biosynthesis from tryptophan (Gluc-Trp) pathway, its pathway genes were enriched 0.5 h after wounding ($P = 0.008$), peaked at 12 h ($P = 0.0008$) and were not significantly enriched by 24 h. In addition, *AT2G38240* (*JASMONATE-INDUCED OXYGENASE4*) and *AT5G05600* (*JASMONATE-INDUCED OXYGENASE2*) from the JA biosynthesis pathway were upregulated at 0.5 h after wounding and remained upregulated throughout the time course (Supplemental Data Set S11). These examples demonstrate the effect of wounding on metabolic pathways and that these wounding-responsive pathways exhibit distinct response patterns. Using the Gluc-Trp pathway as an example, we further assessed the regulatory basis of the wounding responses of genes in this pathway. By 0.5 h after wounding, three Gluc-Trp genes were significantly upregulated and at 1 h three additional genes were significantly upregulated (see stars; Figure 10A). Looking beyond the first hour, we saw a cascading effect, whereby 3 h after wounding, the genes upregulated at 1 h were still upregulated, but the three genes that were first upregulated at 0.5 h were no longer upregulated. Continuing this trend, by 6 h after wounding, only one gene that was upregulated at 1 and 3 h after wounding was still significantly upregulated (Figure 10A). This pattern could be because genes upstream in the pathway are involved, directly or indirectly, in upregulating downstream genes in the pathway.

To understand how the cascading response was regulated, we mapped the pCREs found from each of the wounding-response time-point models built for upregulated genes to the putative promoters of the Gluc-Trp pathway genes (Figure 10B). Starting at 0.5 h after wounding, there was little overlap of important pCREs (red in Figure 10B) across time-points except for pCREs present at 6 and 12 h after wounding. This indicates that for the Gluc-Trp pathway, genes turned on at different times have different CREs. For

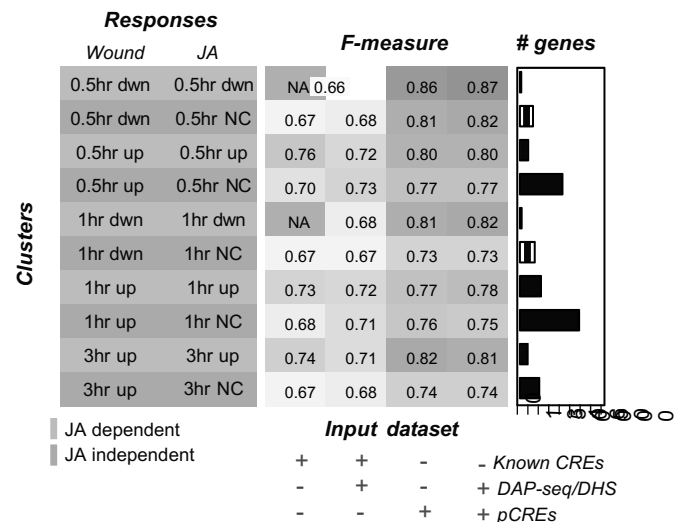


Figure 9 Model performance of each JA-dependent and JA-independent wounding-responsive cluster. Each row is a single cluster (JA-responsive [orange] or JA-non-responsive [blue-NC = no change]), for which a separate model was built using RF. Each column represents the data sets used as features in the model. Known only refers to CREs reported in the literature (Supplemental Data Set S3). DAP-seq and DHS refer to the DAP-seq and DHSs, respectively. FET enriched 6-mer refers to the pCREs, which were enriched for a specific cluster. The F-measure range is from 0.5 (white) to 1 (red), and gradient as well as actual F-measure is shown in each cell. The bar chart next to the heatmap corresponds to each row/cluster and represents the number of genes in that cluster. Note that the models were not generated for genes downregulated 3 h after wounding because there were not enough genes available for training.

example, ACACGT, which resembles bZIP binding motif (PCC = 1), is the most important element at 0.5 h after wounding (Figure 10C) and is not found in Gluc-Trp pathway genes upregulated at other time-points except at 24 h. The treatment-time-specific nature is generally true among the top pCREs except for AACGTG, which was enriched in the promoters of Gluc-Trp pathway genes upregulated 1, 3, 6, 12, and 24 h after wounding. In summary, pCREs responsible for upregulation of Gluc-Trp pathway genes upon wounding varied for different time-points after wounding, indicating that timing of response is an important consideration when identifying CREs. Furthermore, these results highlight that a series of regulatory elements acting at different times, rather than one canonical element, is central to regulating pathways triggered by a specific environment.

Conclusion

The aim of this study was to better understand the temporal differences in transcriptional response to wounding stress in Arabidopsis. We accomplished this by integrating multiple levels of regulatory information (e.g. sequence-based and epigenetic features) into machine learning models of the regulatory code that could be used to predict if a gene was up- or downregulated at a specific time-point after wounding. This system-wide, modeling approach adopted in this

study allows us to address the critical question: how well the known CREs allow for predicting whether a gene would be wound responsive or not. We demonstrated that wounding response is regulated by a diverse set of regulatory elements that are likely bound by TFs from a wide range of TF families, in addition to elements that were previously identified. We identify 4,255 pCREs derived from wounding co-expression clusters which are upregulated at different time-points, with 3,493 (82%) having high sequence similarity (PCC 4 0.8) to known TFBSs, although they are not identical and it is mostly unknown whether they are involved in wound response. These pCREs were more predictive of differential expression at each wounding time-point than models based on known TFBSs (derived from the literature and the DAP-seq database) and information about open chromatin sites. From our machine learning models, we quantified the relative importance of each pCRE included in the model for each time-point. While some pCREs were important across multiple time-points, we generally found that pCREs were either important for early or late time-points

after wounding. Our study also provides a comparison of the *cis*-regulatory programs of JA-dependent and JA-independent responses. We identified 2,569 pCREs important for predicting genes upregulated in response to wounding but not upregulated in response to JA treatment. Of these, 2,371 (92%) had strong sequence similarity (PCC 4 0.8) to known TFBSs. Finally, by focusing on genes in the Gluc-Trp pathway, we identified pCREs important for predicting genes in this wound-responsive specialized metabolite pathway. While our models perform notably better than random expectation, there remains room for improvement. One possible reason we could not predict differential expression more precisely is that we limited our study to focus on CRE sites in the promoter region (+ 1-kb upstream of the transcription start site). However, CREs located in other regions, including the downstream untranslated regions, introns, or coding regions of a gene, can be useful for predicting whether the gene in question is stress responsive (Azodi et al., 2020) and could be evaluated in future studies. Another limitation

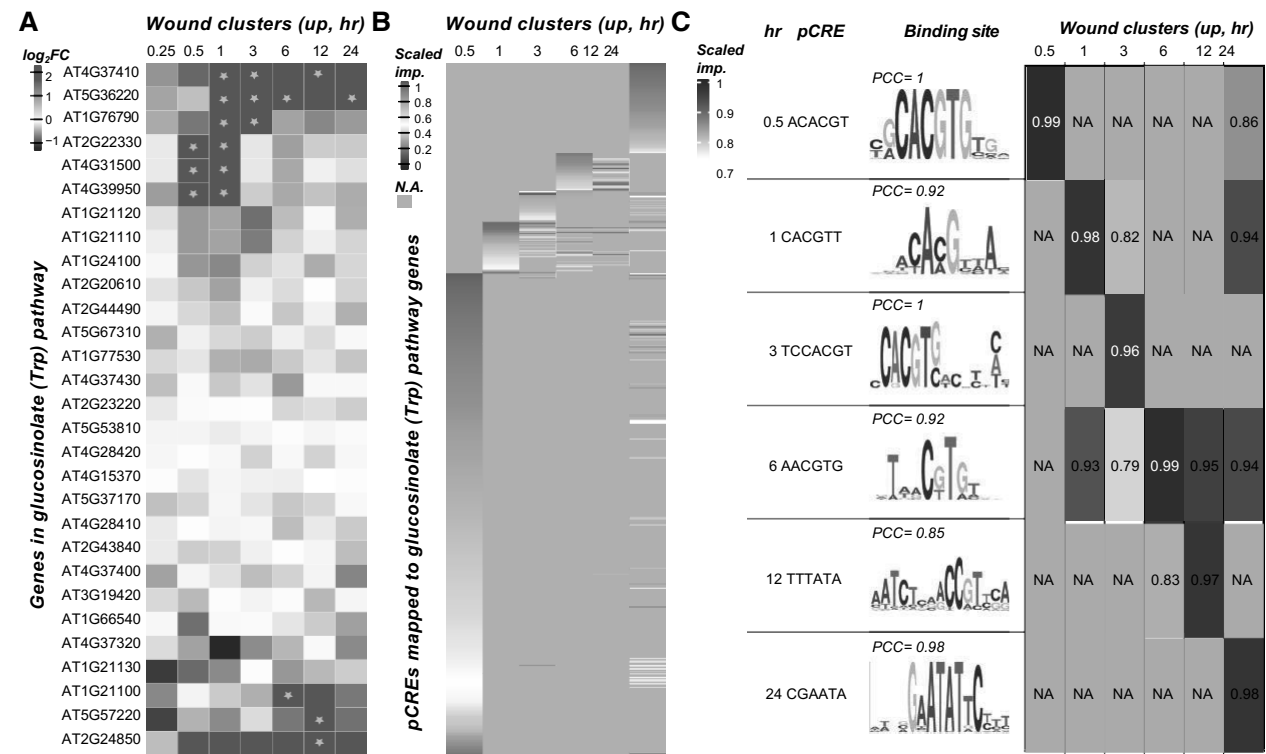


Figure 10 Co-expression and regulation of glucosinolate from tryptophan pathway genes. A, Heatmap showing the log2FC values of all genes in the Gluc-Trp pathway across the seven wounding time-points. Genes are clustered using hierarchical clustering. Genes are on the y-axis, wounding time-points are on the x-axis, and log2FC is represented as the color gradient from a value of 2 or greater (red) to a value of -1 or less (blue). Stars indicate genes are significantly upregulated at a given time-point. B, Scaled importance score of pCREs mapped to Gluc-Trp genes, which are upregulated at a given wounding time-point. Importance is scaled from 0 to 1, where 1 is the most important and 0 is the least important. Each row is a pCRE and each column is the wounding time-point. C, The most important pCRE for Gluc-Trp pathway genes at a given time-point. First three columns show the wounding time-point, pCRE, and the correlation of the pCRE to a known TF binding site shown as a motif logo, respectively. The fourth column shows the scaled importance value of that particular pCRE for the Gluc-Trp genes at each time-point.

is that genes up- or downregulated at a particular time-point might not be all regulated the same way. This is especially likely for larger time-point gene groups, like the 1hr_up cluster, which contained 760 genes. If we could further break down this group, perhaps based on the genes' responses to other stresses, we may be able to model more specific responses at 1 h and improve the overall performance. Finally, the data sets regarding DAP-seq and DHS sites did not come from wounded plants, and therefore were not capturing any changes that may occur in TFBSs or chromatin state after wounding. DAP-seq and DHSs data may indeed improve predictions if drawn from a similar treatment specific data set. Other studies have shown that the association with histone proteins can change under stress response (Kim et al., 2017) and DAP-seq sites have been shown to be important in temporal nitrogen signaling gene expression in Arabidopsis roots and shoots (Varala et al., 2018).

Many of the important pCREs found in this study have not been shown to be associated with wounding. This is especially true for pCREs found at later time-points, which have been less well studied. With technologies such as CRISPR-Cas9, it is feasible to generate precise edits to the DNA to test the role of these pCREs in temporal wounding response experimentally. We mutated the pCRE CCGCGT and this resulted in a significant decrease in expression of the target gene *GER5* under wounding treatment. Additionally, we show three predicted CREs, GTCGGC, GTCACA, and ACACGT, to bind to their predicted TF *in vitro*, and two of these to control protoplast expression *in vivo*, and can be followed up with experiments using stable transgenic lines to show clear association with wound response. Our study demonstrates the feasibility of modeling temporal response to wounding computationally and, through interpreting the models, identifies a set of important putative *cis*-regulatory targets. Finally, the computational framework in this study can be applied to assess the *cis*-regulatory mechanisms in other contexts. We found regulatory sequences previously unknown but likely to be important to JA dependent or independent responses which should be tested in future experiments. Another example is the Gluc-Trp pathway demonstrating that we were able to identify pCREs regulating wound response at the pathway level. Therefore, we expect that the same approach can be applied generally to any sets of genes commonly regulated in a specific environment, stage of development, tissue/cell type, and timing to generate model supported hypotheses of *cis*-regulation.

Materials and Methods

Expression data sets and analysis

Microarray data from three different AtGenExpress studies were downloaded from TAIR and CEL files were processed using the Affy package (1.68.0; Gautier et al., 2004) in R (4.0.1). The studies included biotic stress (Wilson et al.,

2012), abiotic stress (Kilian et al., 2007; Wilson et al., 2012), and hormone treatment (Goda et al., 2008), where wounding is part of the abiotic stress data set. Arabidopsis plants used in those studies grew under similar conditions and were treated 18 days after germination. Those studies were all part of the AtGenExpress project. Each study had eight different treatments of either different stresses or hormones, resulting in a total of 24 data sets. Samples from each data set were collected after treatment at a range of time-points, including 15 min, 30 min, 1 h, 2 h, 3 h, 4 h, 6 h, 12 h, and 24 h after treatment. Note that not all time-points were used in this study for each treatment. For each data set, controls were collected at the same time in order to control for circadian effects.

Differential expression was calculated using Affy and limma (3.46.0) packages in R (Gautier et al., 2004; Ritchie et al., 2015), and significantly differentially expressed genes were those that had an absolute $\log_2FC \leq 1$ ($\log_2FC \leq 1$: upregulated, \log_2FC less than or equal to -1 : downregulated) and adjusted (false discovery rate corrected for multiple comparisons) $P \leq 0.05$. For each treatment within each expression data set, PCC was calculated for all pairwise combinations.

Gene clusters

The wounding time-point clusters were determined by two considerations: (1) time-point after wounding (0.25, 0.5, 1, 3, 6, 12, and 24 h), (2) direction of differential expression (up- or downregulated). For example, genes that were upregulated at 0.25 h after wounding belong to the 0.25hr_up cluster, while genes that are downregulated at 1 h after wounding made the cluster 1hr_down. This created a total of 14 wounding clusters. For wounding and JA clusters, genes were placed in a cluster based on whether they were differentially expressed in one or both treatments at the same time-point. For example, a gene X upregulated in both 1 h after wounding and 1 h after JA treatment would be placed in cluster 1hr_up/1hr_up, while gene Y upregulated in 1 h after wounding but not differentially expressed under 1 h after JA treatment would be placed in cluster 1hr_up/1hr_NC (NC: no changes). Thus, a gene Z upregulated in 1 h after wounding but downregulated in 1 h after JA treatment would be placed in cluster 1hr_up/1hr_down. Therefore, for the three time-points available for both wounding and JA treatment data sets (0.5, 1, and 3 h), there are potentially 18 clusters: 3 time-points \times 2 regulation directions after wounding (up- or downregulated) \times 3 regulation directions after JA treatment (up- or downregulated, or no changes). Three of these potential clusters contained no genes and were subsequently omitted (0.5hr_up/0.5hr_down, 0.5hr_up/1hr_down, 0.5hr_up/3hr_down). In addition, a nondifferentially expressed cluster was determined by genes, which were not differentially expressed across all stress and hormone treatments, including all time-points. For the information of all gene clusters (genes and the number of genes for each cluster) and the overlap between clusters, see Supplemental Data Set S1.

Known CRE curation and pCRE finding

Known regulatory elements, including elements reported to be responsive to JA treatment, wounding, or insect stress, were curated from a literature search (Supplemental Data Set S3). Both the known CREs with experimental evidence and predicted by computational approaches were included. For pCRE finding, putative promoter regions of each gene (identified as 1-kb upstream of the transcription start site) were downloaded from TAIR for Arabidopsis. Homemade python scripts (<https://github.com/ShiuLab/MotifDiscovery>) were used to identify all possible combinations of k -mers (oligomer sequences of length k) present in gene promoters. The Fisher's exact test (FET) was then used to determine the overrepresented pCREs in the promoter region (defined as 1,000-bp upstream of gene start site) for a given wound-responsive gene cluster compared with the nondifferentially expressed cluster. Four P -value cutoffs (adjusted $P \leq 0.01$, $P \leq 0.01$, adjusted $P \leq 0.05$, and $P \leq 0.05$) were explored for the FET, and the second one ($P \leq 0.01$) performed best for the later machine-learning models. Starting with all possible 6-mers, pCREs which were found to be significantly overrepresented in the target clusters were kept. Next, the k -mer finding was performed for 7-mers, which were produced by adding one nucleotide to the enriched 6-mers on either side, thus there were eight possible 7-mers for each 6-mer. These 7-mers were again tested to see if they were significantly overrepresented in the given cluster, and if their P -value was lower than that of the parent 6-mer. If this was true, the 7-mer was kept and the 6-mer was discarded. If not, the 7-mer was discarded and the 6-mer was kept. This progressive procedure of "growing" k -mers continued until the longest k -mer with a P -value lower than its predecessor was obtained. The enriched pCREs were then used as features (present or absent for a pCRE in a gene) to predict whether a gene belongs to a particular wound-responsive cluster or the non-differentially expressed cluster in machine-learning models.

Arabidopsis cistrome and epicistrome

Two data sets with in vitro TFBMs were used to correlate the pCREs with the TFBMs. For the first, the position weight matrices for Arabidopsis TFBMs determined from protein binding arrays (Weirauch et al., 2014) were downloaded from the CisBP database (<http://cisbp.ccbr.utoronto.ca>). For the second, DNA affinity purification sequencing (DAP-seq) peaks (O'Malley et al., 2016) were downloaded from the PlantCistrome Database (<http://neomorph.salk.edu/PlantCistromeDB>). The coordinates of the peaks (which are determined by the previous research group) were then mapped to the Arabidopsis genome using python scripts. If the peak overlapped with the promoter of a gene of interest, the peak was considered present as a feature for that gene. To provide insight into chromatin structure, bed files for DHSs in Arabidopsis (Zhang et al., 2012) were obtained from the National Center for Biotechnology Information database under the ID number GSE34318. The DHS sites were assessed using samples from leaf and flower of both

WT and *ddm1-2* mutant plants. The DHS peak coordinates were obtained from BED files (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE34318>) and then mapped to the Arabidopsis genome. If the peak overlapped with the promoter of a gene of interest, the peak was considered present as a feature for that gene.

Machine learning models

Prediction models were built for each wound-responsive cluster as well as for wounding-JA cluster, where the presence or absence of enriched pCREs from the promoter analysis were used as features to predict whether a gene belongs to the cluster in question or the non-differentially expressed cluster. Two machine learning algorithms implemented in the scikit-learn package (Pedregosa et al., 2011), RF and SVM, were used to build the model for each cluster. Python scripts used to run the models can be found here: <https://github.com/ShiuLab/ML-Pipeline>. For each model, 10% of the data were withheld from training as an independent, testing set. Because the data set was unbalanced (e.g. there were 760 genes in the 1hr_up cluster while 6,855 genes in the nondifferentially expressed cluster [null genes]), 100 balanced data sets were created by randomly drawing genes from the null gene cluster to match with the number of genes in the target cluster. Using the training data, grid searches over the parameter space of RF (max_depth=3, 5, 10, max_features=0.1, 0.5, sqrt, log2, None, n_estimators=100, 500, 1,000) and SVM (Kernel=Linear, C=0.001, 0.01, 0.1, 0.5, 1, 10, 50) were performed. The optimal hyperparameters identified from the grid search were used to conduct a 10-fold cross-validation run (90% of the training data set were used to build the model, the remaining 10% were used for validation) for each of the 100 balanced data sets.

We compare model performance using F-measure defined as:

$$F_1 = \frac{2}{\frac{1}{\text{precision}} + \frac{1}{\text{recall}}} = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}$$

Where Precision = $\frac{tp}{tp+fp}$ and Recall = $\frac{tp}{tp+fn}$ and tp = true positive, fp = false positive, fn = false negative. Throughout the manuscript, we compare the RF models as model performance can change based on the algorithm, but all F-measures are reported in Supplemental Data Set S4. Thus, in a binary model, a perfect prediction has an F-measure of 1 and the F-measure of random expectation is 0.5. The RF models also provide an importance score for each input feature, which is determined by the average decrease in impurity of a node in a decision tree across the forest when the feature is used. Thus, features with higher importance scores are more important for a RF model (Breiman, 2001; Louppe, 2014). Importance values were then normalized by scaling between 0 and 1. Features were ranked based on their importance scores for a model, and the average rank of a feature across five duplicate models run for the same time-point was used as the average importance rank of the

feature. Percentile rank is calculated as dividing the rank of a feature by the total number of features. Performance of regression models were measured using Pearson's correlation of predicted log₂FC to actual log₂FC, where a PCC of 1 is a perfect correlation, 0 indicates no correlation, and -1 indicates an anti-correlation.

For each cluster, models with only known CREs were built, and then models with known CREs plus DAP-seq and DHSs information were built. Finally, models with DAP-seq, DHSs and enriched pCRE information were built. Additionally, for the final model type, five separate models for each wounding time-point cluster, each with 100 balanced replicates were run to determine the rank for each feature (pCREs) from most important to least important. This was then used to get an average importance rank of features from the five models (for average importance ranks, see Supplemental Data Set S7; for raw importance scores from each of the five models, see Supplemental Data Set S8). Before ranking, reverse complement pCREs were removed, so that essentially the same pCRE was not ranked twice. To assess random expectation, gene clusters chosen randomly from the expression data sets, pCREs were found using the same methods as above, and were used to build machine learning models using the methods above. Random gene clusters were made for genes at $n = 30, 50, 100, 150, 200$, and 250 at 20 repetitions each. Model results are reported in Supplemental Data Set S4.

Sequence similarity between pCREs with known TFBSs

TAMO/1.0 (Gordon et al., 2005) was also used to create a tamo file for each pCRE, which was then used to measure the similarity of the pCRE to known TFBSs. To compare pCREs to known TFBSs, pairwise PCC distance between pCREs and TFBSs (both DAP-seq and TFBMs from CIS-bp) was generated using the TAMO program (version 1.0; Gordon et al., 2005). After calculating the PCC distance to all possible TFBSs, the TFBSs with the lowest distance (highest PCC) was determined for each pCRE as its best match and was then used for visualization of the binding site logo. Code for parsing TAMO output can be found at: https://github.com/ShiuLab/MotifDiscovery/tree/master/TAMO_scripts.

CRISPR-Cas9 mutagenesis

The and *CRISPR/Cas9* mutants generated in this study were grown on soil (Suremix growth medium, Michigan Grower Products Inc., USA) or Murashige and Skoog (MS) media (PhytoTech Labs, USA) containing 0.8% agar under a photo-period of 16-h white light/8-h dark at 23°C, with the light provided by fluorescent bulbs of ~100 mmol m⁻² s⁻¹. Wound treatments were done on plants grown on MS medium when they were 18 days old as in Kilian et al. (2007). For each set of three biological replicates, three individual plants were wounded using hemostats as in Koo et al. (2009), plant tissues were pooled and harvested 1 h after wounding, frozen in liquid nitrogen, and then stored at

-80°C until RNA was extracted. For the construction of CRISPR plasmids, two gRNAs were simultaneously assembled into the pHEE401E vector by the Golden Gate assembly method (Wang et al., 2015). The gRNA sequences used in this study are shown in Supplemental Data Set S12. The CRISPR plasmids were transformed into GV3101 *Agrobacterium* strain, followed by floral dipping into *Arabidopsis* (*Col-0*). The T₁ transgenic plants were grown in MS media containing hygromycin (25 mg L⁻¹) for 3 weeks. Genomic DNA was extracted from the rosette leaf of the hygromycin-resistant T₁ plants, and the promoter regions of *JAZ2* and *GER5* were amplified by genomic PCR using the primers, which are listed in Supplemental Data Set S12. The sequences of the regions targeted by CRISPR/Cas9 were validated by Sanger sequencing. All statistical tests for experiments are described in Supplemental Data Set S13.

The expression levels of *JAZ2* and *GER5* were analyzed in the T₂ generation with three biological replicates, for which different sets of seedlings were individually collected. Total RNA from CRISPR-Cas9 mutants was extracted with RNeasy Plant Mini Kit (Qiagen, USA) following manufacturer's instructions. Approximately 500 ng of RNA was used for cDNA synthesis with SuperScript II Reverse Transcriptase (Invitrogen, USA). The transcript levels of *JAZ2* and *GER5* were determined by quantitative real-time polymerase chain reaction (PCR) (Quantstudio 3 Real-Time PCR, Thermo Scientific, USA) using SYBR Green PCR Master Mix followed by manufacturer's instruction (ThermoFisher Scientific, CA, USA). The *C_t* values of the genes were normalized to those of *ACTIN2*. The PCR primer sets were described in Supplemental Data Set S12.

Protein expression and purification

For the DNA-binding affinity test, Gateway donor vectors obtained from the ABRC (Supplemental Data Set S8) were recombined into pDEST17 vector (ThermoFisher Scientific, USA) using LR Clonase (ThermoFisher Scientific, USA). The pDEST17 constructs harboring the TFs were used for the protein expression and purification. For protein purification, the His₆-tagged TFs were transformed into *Escherichia coli* BL21 (DE3) strain. The cells were cultured in 50 mL until reached optical density (OD)₆₀₀ ~0.4 at 37°C. For recombinant protein induction, 0.5-mM isopropyl β-D-1-thiogalactopyranoside was added to the cultured media and incubated at 37°C for 2 h. The cells were collected, resuspended in 5-mL modified phosphate-buffered saline (PBS) buffer (500-mM NaCl, 10-mM Na₂HPO₄, 2-mM KH₂PO₄, 0.05%, and Triton X-100), and lysed by a sonication (Misonix Ultrasonic Liquid Processors S-4000, USA). After centrifugation at 3,500g for 20 min at 4°C, the supernatant was incubated with 100 mL of 50% (w/v) slurry of Ni-NTA agarose (Thermo Fisher Scientific, USA) for 1 h and washed with 10 mL of PBS buffer containing 50-mM imidazole. The resin was eluted by 200 mL of 400-mM imidazole in PBS buffer four times. All protein purification procedures were performed at 4°C. The quality and quantity of protein was verified on sodium dodecyl sulfate-polyacrylamide gel

electrophoresis (SDS–PAGE) gel (15%, 37.5:1 acrylamide:bis-acrylamide, BioRad, USA) followed by Coomassie Brilliant Blue (G-250, Thermo Scientific, USA) staining.

Evaluating DNA-binding activity using ALPHA

DNA binding affinity test was carried out by ALPHA according to the manufacturer's protocol (PerkinElmer, USA; Lee et al., 2021). Briefly, the purified His₆-tagged proteins (0, 50, and 100 nM) were incubated with 0 and 10 nM of streptavidin-conjugated DNA probes for 1 h at room temperature. The mixture of protein and probe was incubated with anti-His₆ AlphaLISA Acceptor beads (20 Ig·mL⁻¹, PerkinElmer, USA) for 1 h at room temperature followed by incubation with AlphaScreen Streptavidin Donor beads (20 Ig·mL⁻¹, PerkinElmer, USA) for 30 min at room temperature. The total mixture was transferred into white 384-well OptiPlate (PerkinElmer, Waltham, MA, USA) and the signal was read in an Alpha-compatible reader. For two biological replicates, two sets of recombinant proteins were obtained from different batches and were subjected to the assays with three technical replicates for each Alpha reaction.

Measuring induced gene expression by RT-qPCR in protoplast

For a luciferase assay, the DNA fragments containing four copies of *cis*-regulatory motifs and corresponding mutant probes were synthesized by Integrated DNA Technologies and cloned into the pENTR/SD/D-TOPO vector (ThermoFisher Scientific, USA). The pENT constructs were subcloned into the pLUC2 vector (Kim and Somers, 2010). The primers and DNA probes used in this study are in Supplemental Data Set S11. Protoplasts were isolated from 3-week-old Col-0 as in a previous study (Arai et al., 2019) with around 1×10^5 cells for each transformation. The isolated protoplasts were co-transfected with 500 ng of reporter construct and 100 ng of Renilla construct containing Renilla luciferase gene driven by 35S promoter (Elomaa et al., 1998) as described previously (Arai et al., 2019). After co-transfection, protoplasts were incubated for 12 h at room temperature in darkness and the luciferase activity was measured by dual-luciferase reporter assay kit (Promega, USA) according to the manufacturer's protocol using a microplate luminometer. Total RNA was extracted from the protoplasts using RNeasy Plant Mini Kit (Qiagen, USA) following manufacturer's instructions. Around 10 ng of RNA was used for cDNA synthesis with SuperScript II Reverse Transcriptase (Invitrogen, USA). The transcript abundance of the gene of interests was determined by quantitative real-time PCR (Quantstudio 3 Real-Time PCR, Thermo Scientific, USA) using SYBR Green PCR Master Mix followed by manufacturer's instruction (ThermoFisher Scientific, CA, USA) by normalizing it to the level of *ACTIN*. The primers used in this study are described in Supplemental Data Set S12.

Pathway enrichment and pCRE mapping

Pathway annotations were downloaded from the Plant Metabolic Network Database (<https://www.plantcyc.org/>). Enrichment tests were performed by using python scripts

(<https://github.com/ShiuLab/GO-term-enrichment>) and the Python Fisher 0.1.9 package, which implements the FET. To map the enriched pCREs to the promoter regions of genes in the glucosinolate from tryptophan (Gluc-Trp) pathway, gff files were created that contained the coordinates of pCREs in the promoters of all Arabidopsis genes. Genes that were annotated in the Gluc-Trp pathway and expressed at a wounding time-point were examined for the presence/absence of the enriched pCREs. Finally, the importance scores of pCREs, which were mapped to Gluc-Trp genes were determined for each wounding time-point model.

Accession numbers

AtGenExpress data from TAIR: <https://www.arabidopsis.org/portals/expression/microarray/ATGenExpress.jsp>

CisBP database: <http://cisbp.cabr.utoronto.ca>

Plant Cistrome Database: <http://neomorph.salk.edu/PlantCistromeDB>

DHS sites: <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE34318>

The following genes and promoter elements can be found in the indicated data sets:

JAZ2 (AT1G74950, Supplemental Data Set S12), *GER5* (AT5G13200, Supplemental Data Sets S9 and S12), *RIBONUCLEASE 1* (AT2G02990, Supplemental Data Set S9), *SULFOTRANSFERASE 2A* (AT5G07010, Supplemental Data Set S9).

Supplemental data

The following materials are available in the online version of this article.

Supplemental Figure 1. Heatmap of the F-measure for all wounding SVM models (supports Figure 2).

Supplemental Figure 2. Mutation in the CACGTG motif of the *JAZ2* promoter led to the downregulation of *JAZ2* expression following wound treatment (supports Figure 6).

Supplemental Figure 3. Recombinant proteins used in ALPHA experiments and ALPHA assay for CRE–TF pairs with no significant binding (supports Figure 7).

Supplemental Figure 4. Wound-responsive expression of the genes downstream of GTCACA or GTCGGC cis-regulatory motifs (supports Figure 8).

Supplemental Figure 5. Gene overlap of JA-dependent and JA-independent clusters (supports Figure 9).

Supplemental Figure 6. Average importance rank for the top 10 pCREs for each JA-dependent and JA-independent wound-responsive model and the associated TF families (supports Figure 9).

Supplemental Data Set S1. Sample cluster overlap and genes in each cluster.

Supplemental Data Set S2. Between sample PCC results.

Supplemental Data Set S3. Known cis-regulatory elements derived from literature.

Supplemental Data Set S4. All machine learning model results.

Supplemental Data Set S5. Feature importance for models using only known elements or sites.

Supplemental Data Set S6. All pCREs enriched for each wounding time-point cluster and their *P*-values.

Supplemental Data Set S7. Summary table for the importance rank of each pCRE for each cluster and their correlation to DAP-seq or cis-BP sites.

Supplemental Data Set S8. Raw importance scores for wounding models.

Supplemental Data Set S9. DNA binding activity of six TFs.

Supplemental Data Set S10. Overall feature importance score for wounding JA-dependent and JA-independent clusters.

Supplemental Data Set S11. Pathway enrichment for each wounding time-point cluster and *P*-values.

Supplemental Data Set S12. Primers used for CRISPR-cas9 and qPCR and promoter sequences of experimental genes.

Supplemental Data Set S13. Statistical analyses for each experimental figure.

Acknowledgments

We thank Sahra Uygun and Nicholas Panchy for helpful discussions and other help that made this work possible.

Funding

This work was partly supported by the National Science Foundation IOS-1733633 to E.G., IOS-1546617, DEB-1655386, DGE-1828149 to S.-H.S., as well as the US Department of Energy Great Lakes Bioenergy Research Center BER DE-SC0018409 to S.-H.S.

Conflict of interest statement. None declared.

References

- Arai H, Yanagiura K, Toyama Y, Morohashi K (2019) Genome-wide analysis of MpBHLH12, a IIIF basic helix-loop-helix transcription factor of *Marchantia polymorpha*. *J Plant Res* 132: 197–209
- Asensi-Fabado M-A, Amtmann A, Perrella G (2017) Plant responses to abiotic stress: the chromatin context of transcriptional regulation. *Biochim Biophys Acta BBA - Gene Regul Mech* 1860: 106–122
- Azodi CB, Lloyd JP, Shiu S-H (2020) The cis-regulatory codes of response to combined heat and drought stress in *Arabidopsis thaliana*. *NAR Genomics Bioinformatics* 2: lqaa049
- Benn G, Wang C-Q, Hicks DR, Stein J, Guthrie C, Dehesh K (2014) A key general stress response motif is regulated non-uniformly by CAMTA transcription factors. *Plant J* 80: 82–92
- Bömer M, O'Brien JA, Pérez-Salamó I, Krasauskas J, Finch P, Briones A, Daudi A, Souda P, Tsui T-L, Whitelegge JP, et al. (2018) COI1-dependent jasmonate signalling affects growth, metabolite production and cell wall protein composition in *Arabidopsis*. *Ann Bot* 122: 1117–1129
- Bostock RM, Pye MF, Roubtsova TV (2014) Predisposition in plant disease: exploiting the nexus in abiotic and biotic stress perception and response. *Annu Rev Phytopathol* 52: 517–549
- Breiman L (2001) Random forests. *Mach Learn* 45: 5–32
- Chung HS, Koo AJK, Gao X, Jayanty S, Thines B, Jones AD, Howe GA (2008) Regulation and function of *Arabidopsis JASMONATE ZIM*-domain genes in response to wounding and herbivory. *Plant Physiol* 146: 952–964
- Delessert C, Wilson I, Van Der Straeten D, Dennis E, Dolferus R (2004) Spatial and temporal analysis of the local response to wounding. *Plant Mol Biol* 55: 165–181
- Doherty CJ, Van Buskirk HA, Myers SJ, Thomashow MF (2009) Roles for *Arabidopsis* CAMTA transcription factors in cold-regulated gene expression and freezing tolerance. *Plant Cell* 21: 972–984
- Dombrecht B, Xue GP, Sprague SJ, Kirkegaard JA, Ross JJ, Reid JB, Fitt GP, Sewelam N, Schenk PM, Manners JM, et al. (2007) MYC2 differentially modulates diverse jasmonate-dependent functions in *Arabidopsis*. *Plant Cell* 19: 2225–2245
- Elomaa P, Mehto M, Kotilainen M, Helariutta Y, Nevalainen L, Teeri TH (1998) A bHLH transcription factor mediates organ, region and flower type specific signals on dihydroflavonol-4-reductase (*dfr*) gene expression in the inflorescence of *Gerbera hybrida* (Asteraceae). *Plant J* 16: 93–99
- Fernández-Calvo P, Chini A, Fernández-Barbero G, Chico J-M, Gimenez-Ibanez S, Geerinck J, Eeckhout D, Schweizer F, Godoy M, Franco-Zorrilla JM, et al. (2011) The *Arabidopsis* bHLH transcription factors MYC3 and MYC4 are targets of JAZ repressors and act additively with MYC2 in the activation of jasmonate responses. *Plant Cell* 23: 701–715
- Figueroa P, Browse J (2012) The *Arabidopsis* JAZ2 promoter contains a G-box and thymidine-rich module that are necessary and sufficient for jasmonate-dependent activation by MYC transcription factors and repression by JAZ proteins. *Plant Cell Physiol* 53: 330–343
- Friggmann H, Gigolashvili T (2014) Update on the role of R2R3-MYBs in the regulation of glucosinolates upon sulfur deficiency. *Front Plant Sci* 5: 626
- Gautier L, Cope L, Bolstad BM, Irizarry RA (2004) affy-analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics* 20: 307–315
- Glisovic T, Bachorik JL, Yong J, Dreyfuss G (2008) RNA-binding proteins and post-transcriptional gene regulation. *FEBS Lett* 582: 1977–1986
- Goda H, Sasaki E, Akiyama K, Maruyama-Nakashita A, Nakabayashi K, Li W, Ogawa M, Yamauchi Y, Preston J, Aoki K, et al. (2008) The AtGenExpress hormone and chemical treatment data set: experimental design, data evaluation, model data analysis and data access. *Plant J* 55: 526–542
- Godoy M, Franco-Zorrilla JM, Pérez-Pérez J, Oliveros JC, Lorenzo Ó, Solano R (2011) Improved protein-binding microarrays for the identification of DNA-binding specificities of transcription factors: protein-binding microarrays for DNA-binding specificities of TFs. *Plant J* 66: 700–711
- Gordon DB, Nekludova L, McCallum S, Fraenkel E (2005) TAMO: a flexible, object-oriented framework for analyzing transcriptional regulation using DNA-sequence motifs. *Bioinformatics* 21: 3164–3165
- van der Graaff E, Laux T, Rensing SA (2009) The WUS homeobox-containing (WOX) protein family. *Genome Biol* 10: 248
- He Y, Gan S (2001) Identical promoter elements are involved in regulation of the OPR1 gene by senescence and jasmonic acid in *Arabidopsis*. *Plant Mol Biol* 47: 595–695
- Heim MA (2003) The basic helix-loop-helix transcription factor family in plants: a genome-wide study of protein structure and functional diversity. *Mol Biol Evol* 20: 735–747
- Higo K, Ugawa Y, Iwamoto M, Korenaga T (1999) Plant cis-acting regulatory DNA elements (PLACE) database: 1999. *Nucleic Acids Res* 27: 297–300
- Howe GA (2004) Jasmonates as signals in the wound response. *J Plant Growth Regul* 23: 223–237

- Howe GA, Jander G (2008) Plant immunity to insect herbivores. *Annu Rev Plant Biol* 59: 41–66
- Hutvagner G, Simard MJ (2008) Argonaute proteins: key players in RNA silencing. *Nat Rev Mol Cell Biol* 9: 22–32
- Ikeuchi M, Iwase A, Ryman B, Lambolez A, Kojima M, Takebayashi Y, Heyman J, Watanabe S, Seo M, De Veylder L, et al. (2017) Wounding triggers callus formation via dynamic hormonal and transcriptional changes. *Plant Physiol* 175: 1158–1174
- Jacobo-Velázquez DA, González-Agüero M, Cisneros-Zevallos L (2015) Cross-talk between signaling pathways: the link between plant secondary metabolite production and wounding stress response. *Sci Rep* 5: 8608
- Kilian J, Whitehead D, Horak J, Wanke D, Weinl S, Batistic O, D'Angelo C, Bornberg-Bauer E, Kudla J, Harter K (2007) The AtGenExpress global stress expression data set: protocols, evaluation and model data analysis of UV-B light, drought and cold stress responses: AtGenExpress global abiotic stress data set. *Plant J* 50: 347–363
- Kim J-M, To TK, Matsui A, Tanoi K, Kobayashi NI, Matsuda F, Habu Y, Ogawa D, Sakamoto T, Matsunaga S, et al. (2017) Acetate-mediated novel survival strategy against drought in plants. *Nat Plants* 3: 1–7
- Kim J, Somers DE (2010) Rapid assessment of gene function in the circadian clock using artificial microRNA in *Arabidopsis* mesophyll protoplasts. *Plant Physiol* 154: 611–621
- Koo AJK, Gao X, Daniel Jones A, Howe GA (2009) A rapid wound signal activates the systemic synthesis of bioactive jasmonates in *Arabidopsis*. *Plant J* 59: 974–986
- LeBrasseur ND, MacIntosh GC, Perez-Amador MA, Saitoh M, Green PJ (2002) Local and systemic wound-induction of RNase and nuclease activities in *Arabidopsis*: RNS1 as a marker for a JA-independent systemic signaling pathway. *Plant J* 29: 393–403
- Lee YS, Herrera-Tequila A, Silwal J, Geiger JH, Grotewold E (2021) A hydrophobic residue stabilizes dimers of regulatory ACT-like domains in plant basic helix–loop–helix transcription factors. *J Biol Chem* 296: 100708
- León J, Rojo E, Sánchez-Serrano JJ (2001) Wound signaling in plants. *J Exp Bot* 52: 1–9
- León J, Rojo E, Titarenko E, Sánchez-Serrano JJ (1998) Jasmonic acid-dependent and -independent wound signal transduction pathways are differentially regulated by Ca^{2+} /calmodulin in *Arabidopsis thaliana*. *Mol Gen Genet* 258: 412–419
- Liu M-J, Seddon AE, Tsai ZT-Y, Major IT, Floer M, Howe GA, Shiu S-H (2015) Determinants of nucleosome positioning and their influence on plant gene expression. *Genome Res* 25: 1182–1195
- Liu M-J, Sugimoto K, Uygun S, Panchy N, Campbell MS, Yandell M, Howe GA, Shiu S-H (2018) Regulatory divergence in wound-responsive gene expression between domesticated and wild tomato. *Plant Cell* 30: 1445–1460
- Lorenzo O, Piqueras R, Sánchez-Serrano JJ, Solano R (2003) ETHYLENE RESPONSE FACTOR1 integrates signals from ethylene and jasmonate pathways in plant defense. *Plant Cell* 15: 165–178
- Loupe G (2014) Understanding random forests: from theory to practice. PhD thesis. University of Liège, Liège, Belgium.
- Ma S, Bohnert HJ (2007) Integration of *Arabidopsis thaliana* stress-related transcript profiles, promoter structures, and cell-specific expression. *Genome Biol* 8: R49
- O'Donnell PJ, Calvert C, Atzorn R, Wasternack C, Leyser HMO, Bowles DJ (1996) Ethylene as a signal mediating the wound response of tomato plants. *Science* 274: 1914–1917
- O'Malley RC, Huang 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
- Omidbakhshfar MA, Proost S, Fujikura U, Mueller-Roeber B (2015) Growth-regulating factors (GRFs): a small transcription factor family with important functions in plant biology. *Mol Plant* 8: 998–1010
- Osakabe Y, Watanabe T, Sugano SS, Ueta R, Ishihara R, Shinozaki K, Osakabe K (2016) Optimization of CRISPR/Cas9 genome editing to modify abiotic stress responses in plants. *Sci Rep* 6: 26685
- Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, et al. (2011) Scikit-learn: machine learning in Python. *J Mach Learn Res* 12: 2825–2830 [Database]
- Reymond P, Weber H, Damond M, Farmer EE (2000) Differential gene expression in response to mechanical wounding and insect feeding in *Arabidopsis*. *Plant Cell* 12: 707–719
- Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK (2015) limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* 43: e47–e47
- Rojo E, León J, Sánchez-Serrano JJ (1999) Cross-talk between signalling pathways determines local versus systemic gene expression in *Arabidopsis thaliana*: Alternative wound signalling pathways in *Arabidopsis*. *Plant J* 20: 135–142
- Rushton PJ, Reinstädler A, Lipka V, Lippok B, Somssich IE (2002) Synthetic plant promoters containing defined regulatory elements provide novel insights into pathogen- and wound-induced signalling. *Plant Cell* 14: 749–762
- Schilmiller AL, Howe GA (2005) Systemic signaling in the wound response. *Curr Opin Plant Biol* 8: 369–377
- Stotz HU, Mueller S, Zoeller M, Mueller MJ, Berger S (2013) TGA transcription factors and jasmonate-independent COI1 signalling regulate specific plant responses to reactive oxylipins. *J Exp Bot* 64: 963–975
- Tierens KFM-J, Thomma BPHJ, Brouwer M, Schmidt J, Kistner K, Porzel A, Mauch-Mani B, Cammue BPA, Broekaert WF (2001) Study of the role of antimicrobial glucosinolate-derived isothiocyanates in resistance of *Arabidopsis* to microbial pathogens. *Plant Physiol* 125: 1688–1699
- Uygun S, Seddon AE, Azodi CB, Shiu S-H (2017) Predictive models of spatial transcriptional response to high salinity. *Plant Physiol* 174: 450–464
- Varala K, Marshall-Colón A, Cirrone J, Brooks MD, Pasquino AV, Léran S, Mittal S, Rock TM, Edwards MB, Kim GJ, et al. (2018) Temporal transcriptional logic of dynamic regulatory networks underlying nitrogen signaling and use in plants. *Proc Natl Acad Sci USA* 115: 6494–6499
- Walley JW, Coughlan S, Hudson ME, Covington MF, Kaspi R, Banu G, Harmer SL, Dehesh K (2007) Mechanical stress induces biotic and abiotic stress responses via a novel cis-element. *PLoS Genet* 3: 13
- Wang Z-P, Xing H-L, Dong L, Zhang H-Y, Han C-Y, Wang X-C, Chen Q-J (2015) Egg cell-specific promoter-controlled CRISPR/Cas9 efficiently generates homozygous mutants for multiple target genes in *Arabidopsis* in a single generation. *Genome Biol* 16: 144
- Weirauch MT, Yang A, Albu M, Cote AG, Montenegro-Montero A, Drewe P, Najafabadi HS, Lambert SA, Mann I, Cook K, et al. (2014) Determination and inference of eukaryotic transcription factor sequence specificity. *Cell* 158: 1431–1443
- Wilson TJ, Lai L, Ban Y, Steven XG (2012) Identification of meta-genes and their interactions through large-scale analysis of *Arabidopsis* gene expression data. *BMC Genomics* 13: 237
- Wittkopp PJ, Kalay G (2012) Cis-regulatory elements: molecular mechanisms and evolutionary processes underlying divergence. *Nat Rev Genet* 13: 59–69
- Wu C-C, Li F-W, Kramer EM (2019) Large-scale phylogenomic analysis suggests three ancient superclades of the WUSCHEL-RELATED HOMEBOX transcription factor family in plants. *PLoS One* 14: e0223521
- Yamamoto YY, Yoshioka Y, Hyakumachi M, Maruyama K, Yamaguchi-Shinozaki K, Tokizawa M, Koyama H (2011) Prediction of transcriptional regulatory elements for plant hormone responses based on microarray data. *BMC Plant Biol* 11: 39
- Yan X, Chen S (2007) Regulation of plant glucosinolate metabolism. *Planta* 226: 1343–1352

- Yang T, Poovaiah BW (2002) A calmodulin-binding/CGCG box DNA-binding protein family involved in multiple signaling pathways in plants. *J Biol Chem* 277: 45049–45058
- Yu C-P, Lin J-J, Li W-H (2016) Positional distribution of transcription factor binding sites in *Arabidopsis thaliana*. *Sci Rep* 6: 25164
- Zhang T, Poudel AN, Jewell JB, Kitaoka N, Staswick P, Matsuura H, Koo AJ (2016) Hormone crosstalk in wound stress response: wound-inducible amidohydrolases can simultaneously regulate jasmonate and auxin homeostasis in *Arabidopsis thaliana*. *J Exp Bot* 67: 2107–2120
- Zhang W, Zhang T, Wu Y, Jiang J (2012) Genome-wide identification of regulatory DNA elements and protein-binding footprints using signatures of open chromatin in *Arabidopsis*. *Plant Cell* 24: 2719–2731