TITLE: Temporal change in chromatin accessibility predicts regulators of nodulation in *Medicago truncatula*

SHORT TITLE: Gene regulation of nodulation in *Medicago truncatula*

AUTHORS

Sara A. Knaack[1]†, Daniel Conde[2]†, Kelly M. Balmant[2], Thomas B. Irving[3], Lucas G. Maia[3], Paolo M. Triozzi[2], Christopher Dervinis[2], Wendell J. Pereira[2], Junko Maeda[3], Sanhita Chakraborty[3], Henry W. Schmidt[2], Jean-Michel Ané[3,4], Matias Kirst[2,5]*, Sushmita Roy[1,6,7]*

AFFILIATIONS

[1] Wisconsin Institute for Discovery, University of Wisconsin, Madison, Wisconsin 53715, WI 53715, USA.

[2] School of Forest Resources and Conservation, University of Florida, Gainesville, FL 32611, USA.

[3] Department of Bacteriology, University of Wisconsin, Madison, WI 53706, USA.

[4] Department of Agronomy, University of Wisconsin, Madison, WI 53706, USA.

[5] Genetics Institute, University of Florida, Gainesville, FL 32611, USA.

[6] Department of Computer Sciences, University of Wisconsin, Madison, WI 53792, USA.

[7] Department of Biostatistics and Medical Informatics, University of Wisconsin, Madison, WI 53792, USA.

† These authors contributed equally to this work

* Corresponding authors: S. Roy (sroy@biostat.wisc.edu), M. Kirst (mkirst@ufl.edu)

ABSTRACT

Rhizobia can establish associations with legumes to provide plants with nitrogen, in a critical symbiosis for agricultural systems. Symbiosis triggers extensive genome and transcriptome remodeling in the plant, yet the extent of chromatin changes and impact on gene expression is unknown. We combined gene regulatory features and their chromatin accessibility (ATAC-seq) profile to predict the temporal transcriptome (RNA-seq) dynamics of roots treated with rhizobia lipo-chitooligosaccharides. Using a novel approach, Dynamic Regulatory Module Networks, we predicted gene expression as a function of regulatory feature chromatin accessibility. This approach uses regularized regression and identifies the *cis*-regulatory elements and associated transcription factors that most significantly contribute to transcriptomic changes triggered by lipo-chitooligosaccharides. Regulators involved in auxin (*SHY2*), ethylene (*EIN3*), and abscisic acid (*ABI5*) hormone response, as well as histone and DNA methylation (*IBM1*), emerged among those most predictive of transcriptome dynamics.

KEYWORDS: nitrogen fixation, nodule, chromatin accessibility, transcriptome and chromatin integration, gene regulatory network, predictive models

MAIN

Legumes can establish a well-characterized mutualism with nitrogen-fixing rhizobia. Signal exchanges between the host plant and bacteria initiate intracellular infection of host cells, followed by the development and colonization of root nodules (1). Nodules provide a unique niche for the bacteria and fix nitrogen, allowing plants to grow with little to no outside sources of nitrogen and even build soil nitrogen levels for subsequent crops (2). Hence, understanding symbiotic processes between legumes and rhizobia is extremely valuable for the productivity and sustainability of agricultural systems worldwide.

Symbiosis begins with compatible rhizobia detecting flavonoids and isoflavonoids produced by the legume host (3), and subsequent release lipo-chitooligosaccharides (LCOs) by the bacteria. The host plant perceives LCOs with LysM domain receptor-like kinases heterodimers, such as Nod Factor Perception (NFP) and LysM Domain Receptor-like Kinase 3 (LYK3) in *Medicago truncatula* (4, 5). LCO perception activates a signaling cascade, including the plasma membrane-localized LRR-type receptor kinase Doesn't Make Infections 2/Nodulation Receptor Kinase (MtDMI2/MtNORK), the calcium-regulated calcium channel (MtDMI1), cyclic nucleotide-gated calcium channels, *M. truncatula* calcium ATPase 8 (MtMCA8), and including components of the nuclear pore complex (6–8). The cascade results in oscillations of nuclear calcium concentrations, detectable by the nucleus-localized Calcium/Calmodulin-dependent protein Kinase (CCaMK, MtDMI3 in *M. truncatula*) (9). CCaMK activates the transcription factor (TF) Interacting Protein of DMI3 (MtIPD3/CYCLOPS). Downstream, other TFs are activated, such as Nodulation-Signaling Pathway 1 and 2 (NSP1 and NSP2), Nodule INception (NIN), Ethylene Response Factor required for Nodulation 1,2 and 3 (ERN1, 2 and 3), and Nuclear Factor YA-1 and YB-1 (NF-YA1 and NF-YB-1) (10, 11).

The coordinated activity of these TFs triggers transcriptional changes (12) essential for infection of the root hair cells (in *M. truncatula*), nodule organogenesis and infection of the nodule cortex (10, 11). These processes require changes in chromatin accessibility (13), i.e., the distribution and organization of nucleosomes. Densely packed nucleosomes limit access of regulators to genomic regions, yet nucleosomes are depleted at active regulatory loci. Nucleosome occupancy is dynamic and creates a continuum of accessibility from closed to open; a critical determinant of cell function (14). In plants, chromatin reorganization regulates photomorphogenesis and flowering (15, 16). Active DNA demethylation by DEMETER (DME) is critical for gene expression reprogramming during nodule differentiation in *M. truncatula*, and the acquisition of organ identity (13). Also, in *M. truncatula*, the gene expression level of Nodule-specific Cysteine-Rich genes (NCR) across root nodule zones are correlated with chromatin accessibility (17).

The extent of chromatin accessibility change and impact on transcriptional regulation in rhizobial infection, colonization, and nodule development, remains unknown. Thus, we measured temporal changes in the transcriptome (RNA-seq) and genome-wide chromatin accessibility (ATAC-seq) in response to *Sinorhizobium meliloti* LCOs in *M. truncatula* roots. To characterize the gene regulatory network, the role of chromatin accessibility, and its consequent impact on the transcriptional dynamics of LCO response, We applied a novel algorithm, Dynamic Regulatory Module Networks (DRMN), to predict gene expression as a function of chromatin accessibility profiles of *cis*-regulatory features. DRMN results suggest that chromatin accessibility plays an essential role in regulating the transcriptional dynamics in response to LCOs and identified the regulatory features (and associated TFs) that contribute most to transcriptional change during symbiosis between rhizobia and *M. truncatula*.

RESULTS

Measurement of transcriptome and chromatin accessibility in response to LCO

Our approach to study global transcriptomic and chromatin accessibility under rhizobium LCO signaling in *M. truncatula* used the genotype Jemalong A17, treated with LCOs purified from *S. meliloti*. Samples were analyzed for the control (t=0 hr) condition, at 15 min, 30 min, and at 1, 2, 4, 8, and 24 hours after treatment. For each time point three biological replicates were characterized with RNA-seq (transcriptome), and one replicate with ATAC-seq (chromatin accessibility). An LCO concentration of 10^{-8} M was used, as in previous studies (18, 19).

Root transcriptome response to LCOs involves genes implicated in *Medicago* response to rhizobia and early nodule development

8.7-17.9 million short sequence reads (2×100 bp) were generated from the 24 RNA-seq libraries, with alignment rates between 87-95%. Principal component analysis (PCA) showed a clustering of biological replicates and a time-dependent variation, with the first component explaining ~30% of data variation (**Fig. 1A**). Comparison of expression levels at each time point (relative to control, t=0 hr) revealed 12,839 differentially expressed genes (DEGs) with significant change in expression (adjusted $P < 0.05$), including 7,540 and 7,051 that were up- and down-regulated in at least one time point relative to control, respectively.

To validate these results the identified DEGs were compared to DEGs identified from previously published time course data in *M. truncatula* roots inoculated with rhizobium (12). Comparisons were made to DEGs in the following genotypes: the Jemalong A17 wild type, the LCO-insensitive *nfp* mutant, the infection *lyk3* mutant, and the LCO-hypersensitive *skl* mutant.

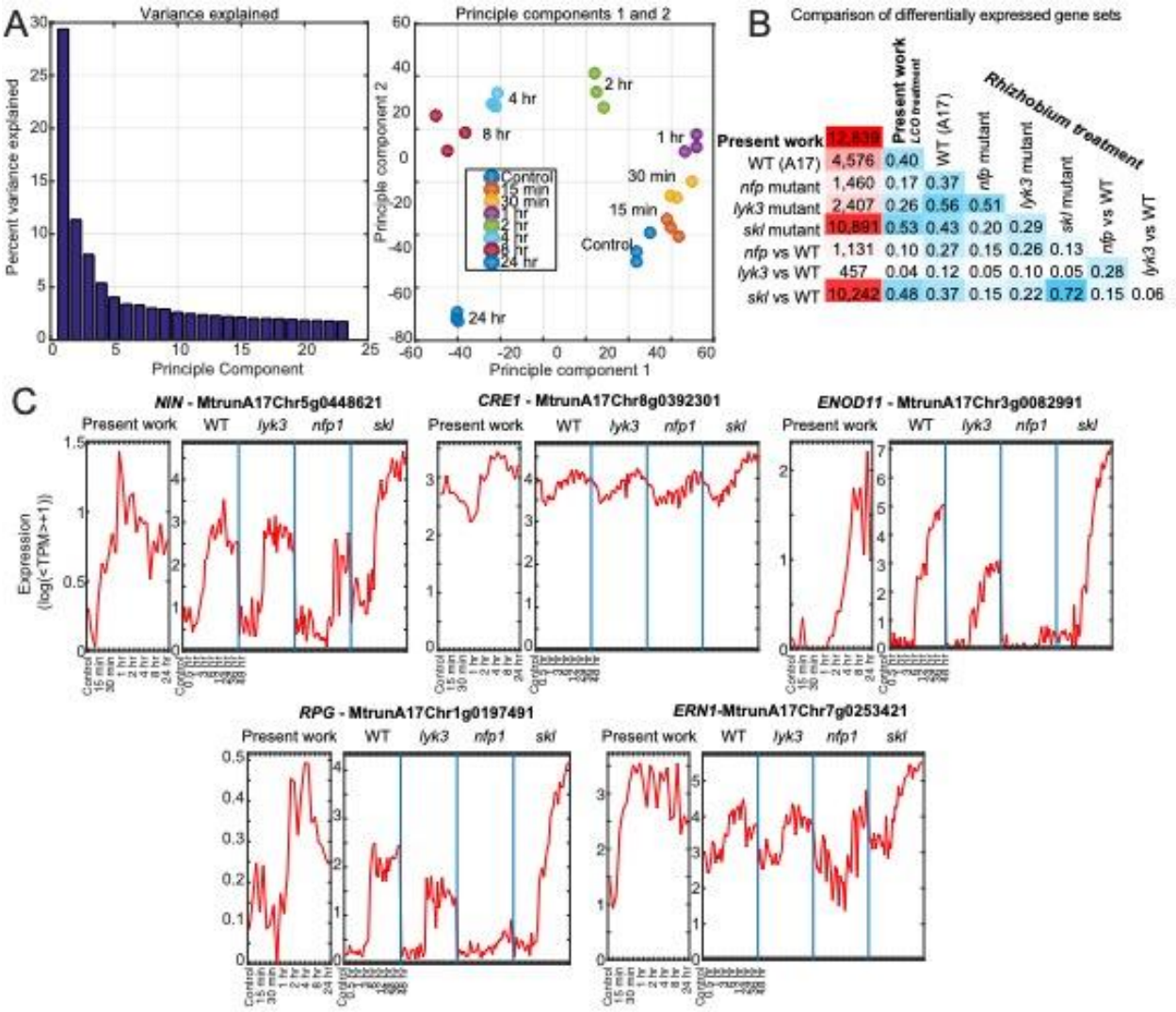


Fig. 1 Transcriptome data summary. (A) PCA presents partitioning and time ordering of biological replicates across time points. Principle components 1, 2, and 3 explain ~50% of the time-dependent variation. **(B)** Similarity scores (f-score) between the DEG set obtained in this study (present work) and DEG sets identified from previously published data (12). **(C)** Expression patterns of *NIN*, *CRE1*, *ENOD11*, *RPG* and *ERN1* across time points, and in the four genotypes analyzed by Larrainzaer et al (12).

The highest similarity, measured by f-score, to our DEG set was observed for the mutant genotype most sensitive to LCOs, *skl* (0.53). In contrast the similarity for DEGs in the Jemalong A17 wild-type strain was 0.40. Expectedly, the similarity was lowest for the *lyk3* (0.26) and *nfp* (0.17) mutants (**Fig. 1B**). The results show similar trends in the DEGs detected in each genotype (12), and indicate LCOs induced the transcriptional changes observed in our experiment. Furthermore, marker genes for rhizobium-induced nodulation were up-regulated (compared to t=0 hr), including *NIN* (*NODULE INCEPTION*, induced after 15 min, with a maximum induction at t=1 hr), *CRE1* (*Cytokinin Response 1*, differentially expressed at 4 hr, 8 hr, and 24 hr), *ENOD11* (*Early nodulin 11*, highly induced at 8 hr and 24 hr), *RPG* (*Rhizobium*-directed polar growth, at 4 hr) and *ERN1* (*Ethylene Responsive Factor Required for Nodulation 1*, induced as early as t=15 min, **Fig. 1C**).

To examine transcriptome dynamics and identify genes involved in LCO response, we applied ESCAROLE, a probabilistic clustering algorithm designed for non-stationary time series (20). The transcriptome data were clustered into seven modules at each time point based on expression levels (very low, low, medium-low, medium, medium-high, high and very high expression) (**Fig. 2A**). The choice of seven modules maximized the log-likelihood and silhouette index (**Fig. S1, A and B**). Next, transitioning genes (genes that change module assignment over time) were identified. Of 37,356 genes 12,261 were transitioning, including several implicated in symbiosis (**Fig. S1C**). Transitioning genes with similar dynamics were clustered using hierarchical clustering, identifying 112 clusters (>=10 genes each) (**Fig. 2B**) including 11,612 genes.

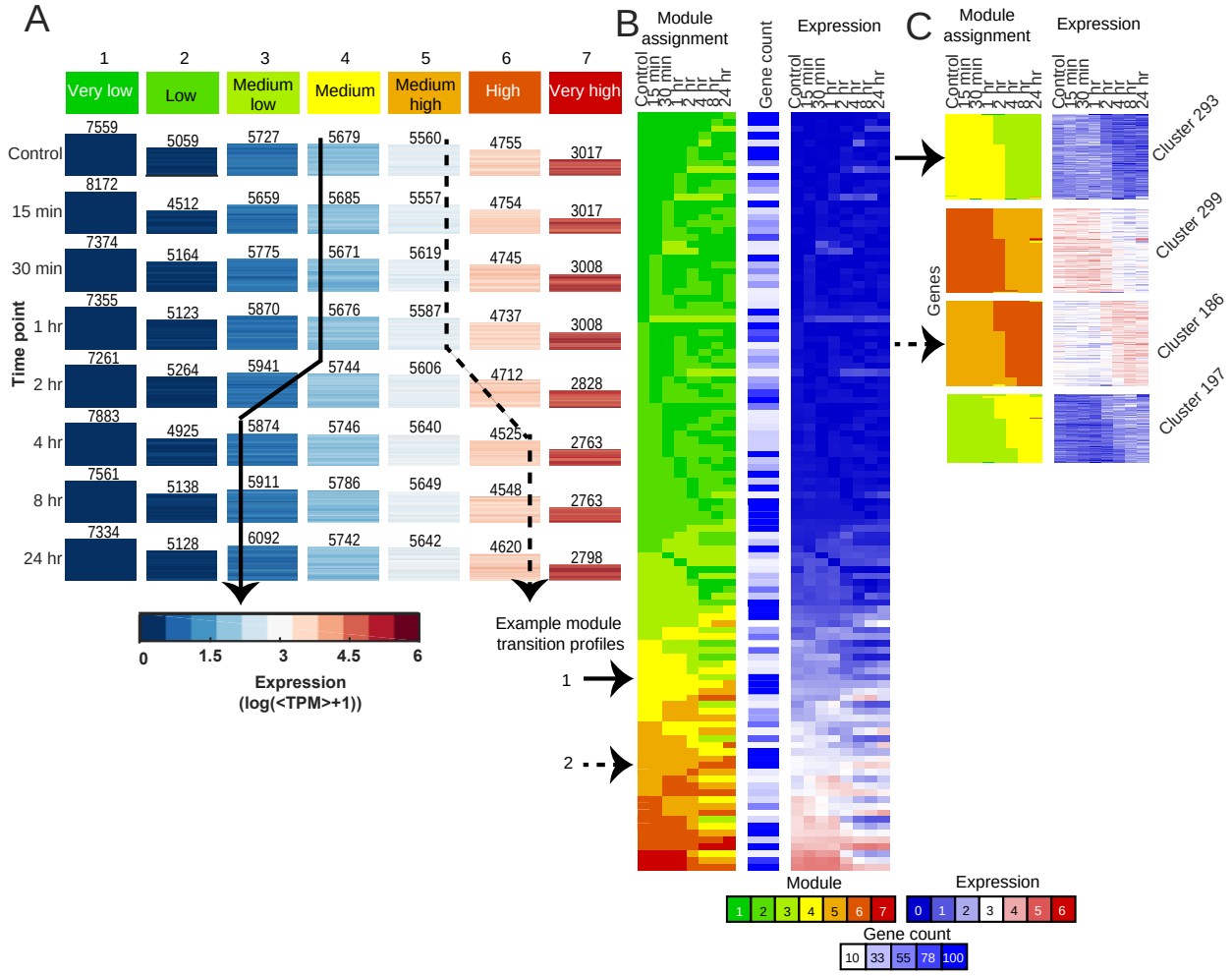


Fig. 2 Transcriptome dynamics in response to LCOs. (A) Expression of genes measured at each experimental time point grouped into seven modules based on transcript abundance. **(B)** The module assignment heatmap depicting typical gene trajectories obtained by hierarchical clustering of gene module profiles into transitioning gene sets. Shown are the mean module assignment at each time point for genes in each set. **(C)** Example of transitioning gene sets showing gene expression upregulation or downregulation, enriched for genes implicated in nodulation such as defense response to bacterium (Cluster 293) or meristem growth (Cluster 186).

The transitioning gene clusters were evaluated for Gene Ontology (GO) category enrichments. Among clusters representing downregulation of expression over time, several were enriched for genes implicated in defense responses to bacterium (cluster 293, downregulated from 2-4 hr), and the biosynthesis of plant hormones involved in the suppression of nodulation (**Fig. 2C**). For instance, cluster 299 (downregulated after 2 hr) is enriched in jasmonic acid (JA) biosynthesis and jasmonic acid response genes, including *CORONATINE INSENSITIVE 1* (*COI1*), which forms part of the JA co-receptor complex for the perception of the JA-signal (21). Among the gene clusters upregulated over time, several are implicated in early stages of symbiosis and nodule development. For instance, cluster 186 (induced 2-4 hr after LCO treatment; **Fig. 2C**) is enriched in genes implicated in the regulation of meristem growth, including an *ARABIDOPSIS TRITHORAX 3* (*ATX3*) homolog (MtrunA17Chr4g0005621) and a *LATERAL ORGAN BOUNDARIES DOMAIN* (*LBD*) transcription factor (MtrunA17Chr4g0043421). *ATX3* encodes an H3K4 methyltransferase (22), and LBD proteins are characterized by a conserved LATERAL ORGAN BOUNDARIES (LOB) domain and are critical regulators of plant organ development (23), including lateral roots and nodules (24). This cluster also contains *EPP1* and the cytokinin receptor *CRE1*, both positive regulators of early nodule symbiosis and development (25, 26). Other essential regulators of LCO signaling, such as *DMI1* (cluster 197, **Fig. 2C**), *NIN* (cluster 205), *NF-YA1* (cluster 177) and the marker of LCO perception *ENOD11* (cluster 296) are also found in clusters exhibiting induction under LCO treatment (**Fig. S1D**). Overall, the ESCAROLE analysis characterized the gene expression dynamics of *M. truncatula* roots in response to the LCOs and recapitulated known features of the molecular cascade associated with this process.

LCO treatment causes genome-wide changes in chromatin accessibility

To study chromatin accessibility changes in response to LCOs, we performed ATAC-seq on samples collected at t=0 hr and each treatment time point, obtaining 54-235 million paired-end reads with alignment rates of 46-75% (**Fig. S2, A** and **B**). We evaluated chromatin accessibility in gene promoter regions, defined as ±2 kbp around the transcription start site (TSS). To quantify chromatin accessibility within promoters we summed the per base pair (per-bp) read coverage within that region, for each time point. For each time point the log-ratio of per-bp read coverage in each promoter was taken relative to the global mean of per-bp coverage and quantile normalized before applying PCA and k-means clustering. We partitioned the resulting promoter accessibility profiles into six characteristic patterns (clusters) using k-means clustering (**Fig. 3A**, **Fig. S3A**). Clusters 1 and 6 present general patterns of decrease and increase in promoter accessibility, respectively, whereas clusters 2-5 present more transient variation. The correlation of accessibility between time points suggests an overall reorganization of promoter accessibility 1-2 hr after the treatment (**Fig. S3B**). The temporal change in accessibility profiles is evident for the promoter regions of several nodulation genes, including *CRE1*, *CYCLOPS*, and *EIN2* (**Fig. 3B, Fig**. **S3**, **C** and **D**). A PCA of the promoter signals showed time-dependent variation (**Fig. 3C**), with the first component explaining >50% of the variance.

We called peaks for each time point using the Model-based Analysis of ChIP-Seq version 2 (MACS2) algorithm (27) (**Fig. S2C**), and subsequently merged peaks from different time points with at least 90% overlap between them into what are referred to hereafter as universal peaks (**Fig. S2D**).

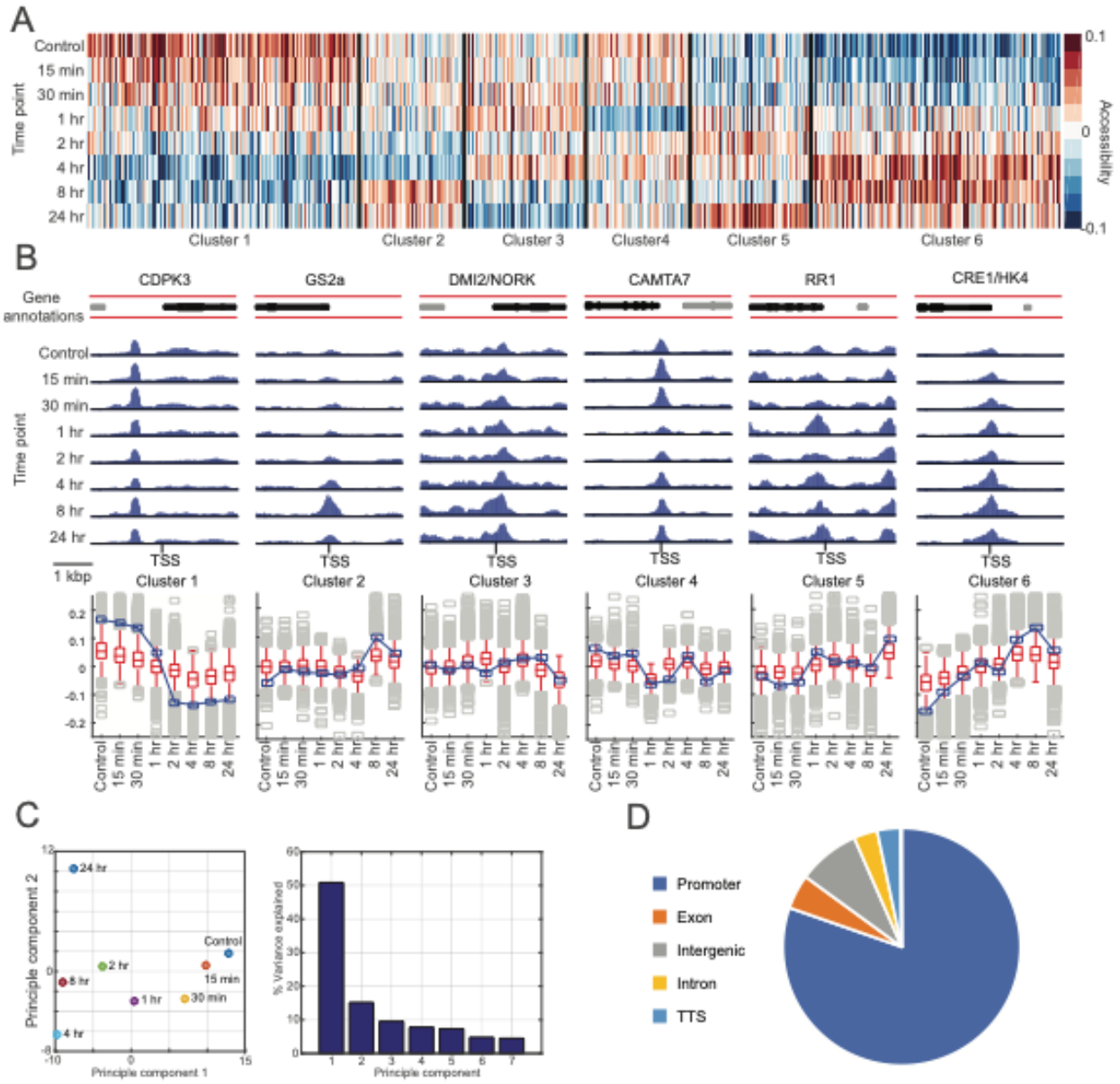


Fig. 3 Chromatin accessibility data exploratory analysis. (A) Clustering of promoter accessibility profiles in the LCO-response time course. **(B)** Plots of the normalized coverage profiles for the promoter regions (± 2 kbp from TSS) of known regulatory genes of root nodulation, representing each cluster (upper panel). In black (top of **B**) are the genes of interest, and in gray, other proximal genes. **(C)** PCA results for the same promoter accessibility data. **(D)** Genomic distribution of chromatin accessibility universal peaks identified by ATAC-seq.

Chromatin accessibility peaks showed a robust genomic distribution across time points, with 81-85% of peaks located within 10 kbp upstream and 1 kbp downstream of a gene TSS (**Fig 3D, Fig. S2, C and E**), spanning 50.4 Mbp (11.7%) of the *M. truncatula* (v5) genome. As with the promoter accessibility clustering, clustering accessibility profiles of universal peaks identified distinct patterns of temporal change (**Fig. S2G**). These results demonstrate LCO treatment had a genome-wide impact on chromatin accessibility, prospectively associated with simultaneous change in gene expression.

Chromatin accessibility is correlated with transcriptional dynamics of nodulation genes

We evaluated the relationship between chromatin accessibility and gene expression using two approaches. First, we correlated promoter (±2 kbp around the TSS) chromatin accessibility and corresponding gene expression profiles. Second, we correlated gene expression with chromatin accessibility profiles of universal peaks within 10 kbp upstream and 1 kbp downstream of a gene TSS.

Our first approach of correlating promoter accessibility and gene expression profiles identified 6,429 genes with significant correlation (**Fig. 4A,** $P<0.05$); 4,777 with positive correlation and 1,652 with negative correlation (**Fig. 4B**), representing 17.2% of the 37,356 genes analyzed. Among these were 36 genes with known roles in symbiosis (**Fig. S3C**), including *ERN1*, *CRE1*, *LYK10/EPR3*, *SKL/EIN2*, and *IDP3*/*CYCLOPS* with positive correlation, and *LYK8*, *ERN2*, *CAMTA3*, and *CAMTA4* with negative correlation. We next examined significantly correlated genes (**Fig. 4A**) in the clusters obtained from promoter accessibility profiles (**Fig. 3A**), and visualized the expression and accessibility profiles ordered by the clusters (separately for positive and negative correlation). This revealed robust patterns of consistency between promoter accessibility and expression (**Fig. 4B**).

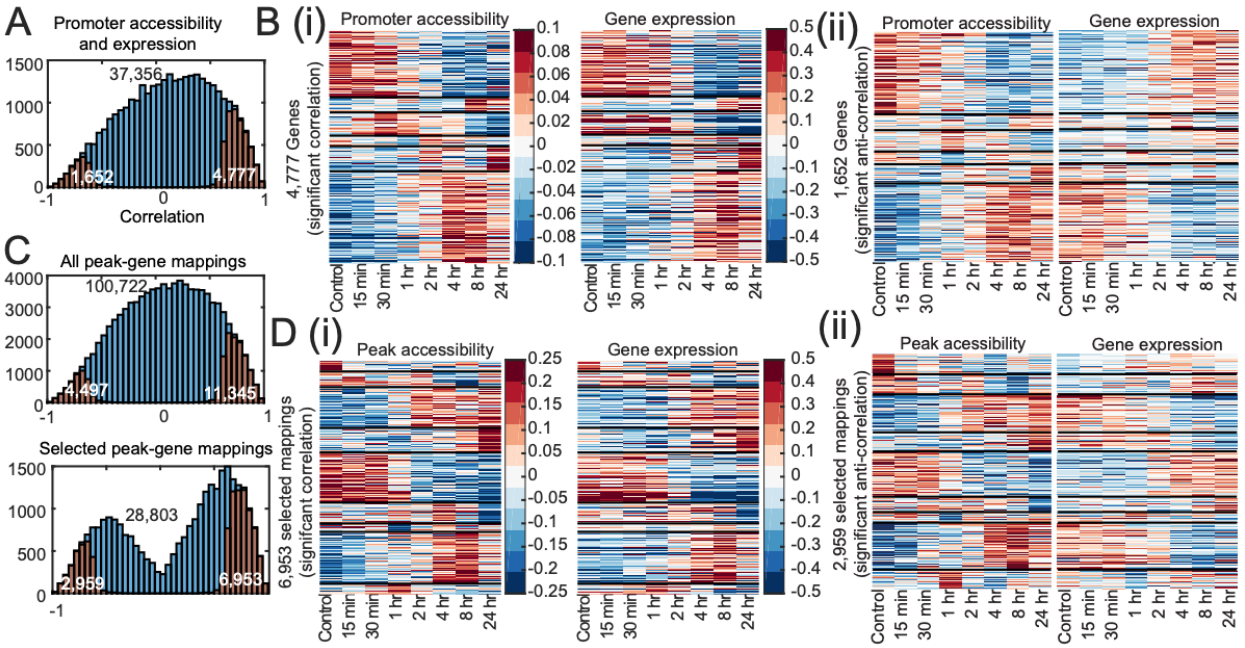


Fig. 4 Correlation between chromatin accessibility and gene expression. (A) Histogram of Pearson correlation of promoter accessibility and gene expression profiles (blue), instances of significant correlation (orange), and statistics (inset numbers). **(B)** Clusters of promoter accessibility and gene expression for significant ($p \leq 0.05$) (i) positive and (ii) negative correlation relative to random. **(C)** Histograms of correlation for all (blue) and significantly correlated peaks and genes (orange), and associated statistics. Shown for all mapped peak-gene pairs (above), and those selected for maximal correlation for each gene (below). **(D)** Clustered peak accessibility profiles and corresponding expression profiles for selected mappings of peaks to genes that are significantly (i) positively and (ii) negatively correlated.

In the second approach, we focused on universal peaks within 10 kbp upstream to 1 kbp downstream of gene TSSs (**Fig. 4C**). Of 125,140 such mappings of peaks to genes, 100,722 were associated with 28,803 (of 37,536) expressed genes. Peak accessibility was significantly correlated with gene expression in 15.7% of these pairings (**Fig. 4C**), comparable to the 17.2% (6,429) genes with significant correlation between expression and promoter accessibility. When considering each gene and only the most correlated peak (28,803 selected pairs), 34.4% (9,912 genes) were significantly correlated, including 56 nodulation genes (**Fig. 4D**).

Finally, the ESCAROLE-defined transitioning gene clusters exhibited coordinated trends between promoter accessibility and gene expression (**Fig. 2B**, **Fig. S1D**). 2,501 of the 11,612 (21.5%) transitioning genes exhibited significant correlation between their profiles of expression and promoter chromatin accessibility. These results suggest that chromatin accessibility is an important regulatory mechanism in transcriptional response to LCOs.

DRMN integration of ATAC-seq and RNA-seq data identifies key regulators that determine gene expression dynamics in response to LCOs

Based on the observation that chromatin accessibility may contribute to transcriptional control of rhizobia-plant symbiosis, we applied a novel Dynamic Regulatory Module Networks (DRMN) (28) approach to analyze the RNA-seq and ATAC-seq time course data. DRMN predicts gene expression as a function of regulatory features (29) by first grouping genes into modules based on expression levels (similar to ESCAROLE), and then learning a regulatory program for each module. DRMN uses regularized regression for the predictive model and leverages the temporal nature of a data set using multi-task learning (30) to simultaneously learn these predictive models for each module in each time point.

We applied DRMN with seven expression states using two types of features (**Fig. 5A**): (1) the aggregated signal of ATAC-seq reads in gene promoters (defined as ± 2 kbp of a gene TSS), and (2) the ATAC-seq signal on genomic coordinates of known motifs within -10 kbp and +1 kbp of a gene TSS. Both feature types represent chromatin accessibility, but the first is independent of the presence of known motifs, whereas the second represents the accessibility of motif sites. Motif features were based on the CisBP v1.2 database for *Medicago* (31), and curated motifs for several known regulators of root nodulation, including CYC-RE, NSP1, NIN and the nitrate response *cis*-element (NRE). Hyper-parameters for DRMN were selected using a grid search with cross-validation error and quality of inferred modules (**Fig. S4A**). Notably, the DRMN modules represent statistically different expression levels (**Fig. S4B,** Kolmogorov-Smirnov-test $P < 10^{-300}$). To assess the predictive power of DRMN we correlated predicted and measured expression levels in three-fold cross-validation tests (**Fig. 5B**). The mean correlation of predicted and measured values per module was 0.26-0.46 (**Fig. S4C**) across all modules and time points. The module with lowest expression was most difficult to predict.

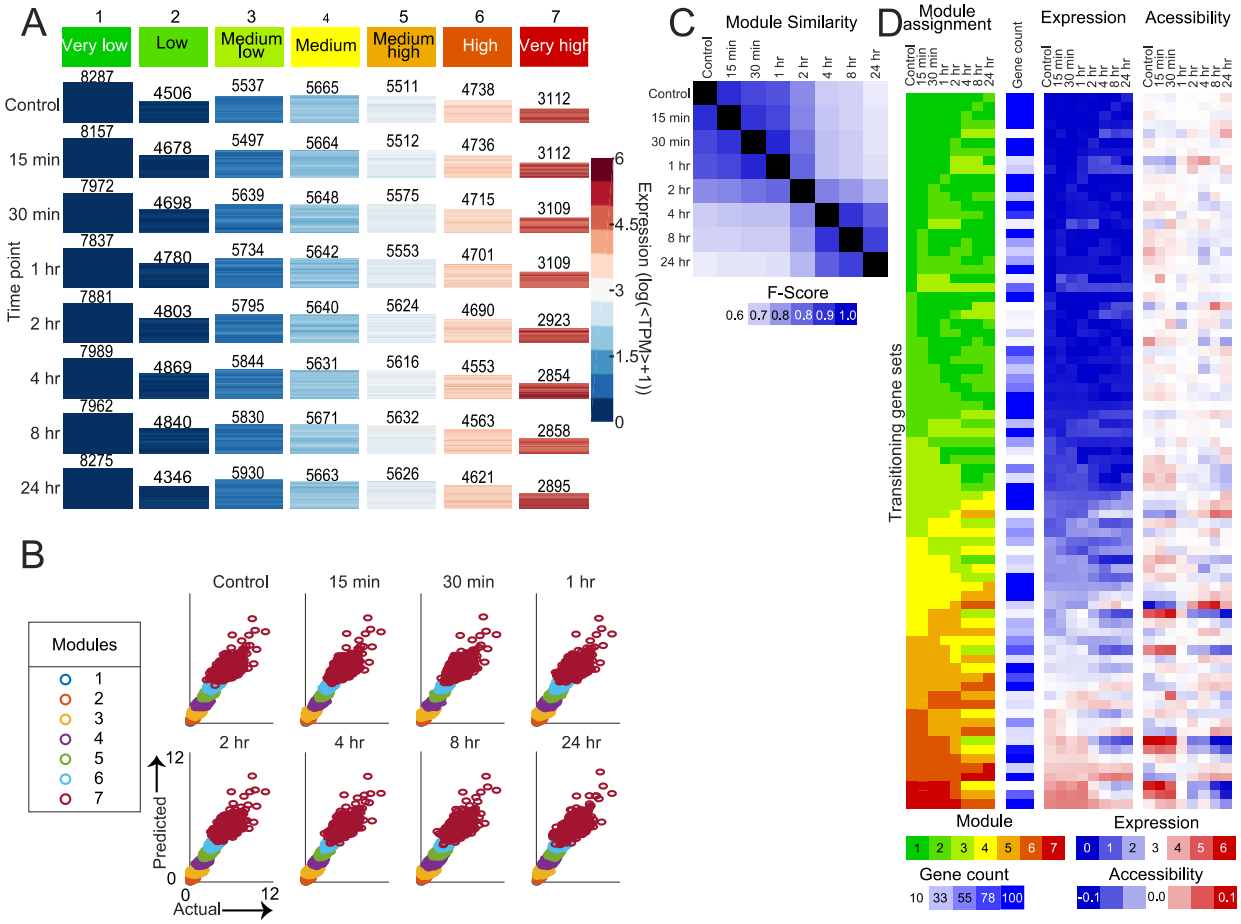
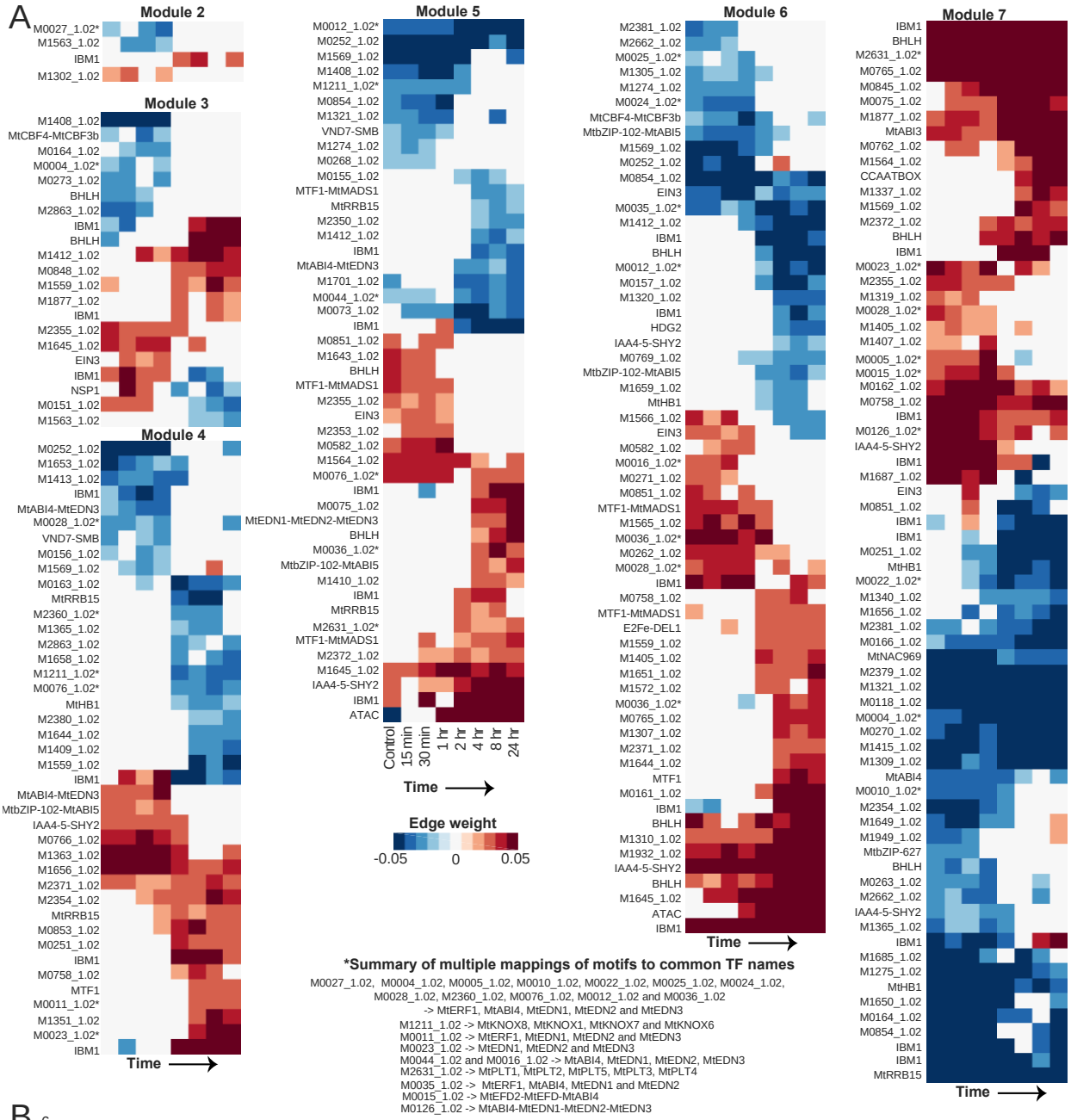


Fig. 5 Dynamic Regulatory Module Network analysis. (A) Heatmap of DRMN inferred expression modules across the time course and **(B)** scatter plots of actual and predicted values. **(C)** F-score similarity of DRMN modules identified at each time point. **(D)** Transitioning gene sets. Shown are the mean module assignment, expression levels and accessibility levels for each gene set (rows) across time (columns).

Next, we compared module gene sets for each pair of time points (**Fig. 5C**), and found that most significant module reorganization occurred at ~2 h. The modules are more similar (F-score 0.88-0.94) before and after 2hr, than across this time point (F-score <=0.70), suggesting significant, coordinated regulatory change. This indicates general reorganization of chromatin accessibility ~1-2 hr after the treatment, likewise supported by the promoter accessibility analysis.

We next obtained sets of transitioning genes that changed expression state across time (**Fig. 5D**), similar to those from ESCAROLE. We identified 79 transitioning gene clusters including 10,176 genes, many of which were differentially expressed (hypergeometric test overlap P <0.05). The overlap of the transitioning gene sets from ESCAROLE and DRMN is 77%, indicating consistency between the two analyses for defining dynamic genes. The DRMN transitioning gene clusters included 43 known regulators of nodulation such as *NIN*, *NF-YA1*, several *LYK* genes (4, 5, 6, and 10), *ENOD8* and *ENOD11*.

We used the DRMN results to prioritize regulators that shape the transcriptional program in response to LCOs. We tested the modules for enrichment of motifs bound by known regulators. We also filtered the regulatory features identified in DRMN models for significant temporal variation (**Fig. 6A**). Specifically, we applied a *t*-test to identify regulators whose coefficient changed significantly between the 0-2 hr vs 4-24 hr portions of the time course (P<0.01), which timeframe was associated with the most extensive reorganization of expression modules (**Fig. 5C**).


A
Module 2
M0027_1.02*
M1563_1.02
IBM1
M1302_1.02
Module 3
M1408_1.02
MtCBF4-MtCBF3b
M0164_1.02
M0004_1.02*
M0273_1.02
BHLH
M2863_1.02
IBM1
BHLH
M1412_1.02
M0848_1.02
M1559_1.02
M1877_1.02
IBM1
M2355_1.02
M1645_1.02
EIN3
IBM1
NSP1
M0151_1.02
M1563_1.02
Module 4
M0252_1.02
M1653_1.02
M1413_1.02
IBM1
MtABI4-MtEDN3
M0028_1.02*
VND7-SMB
M0156_1.02
M1569_1.02
M0163_1.02
MtRRB15
M2360_1.02*
M1365_1.02
M2863_1.02
M1658_1.02
M1211_1.02*
M0076_1.02*
MtHB1
M2380_1.02
M1644_1.02
M1409_1.02
M1559_1.02
IBM1
MtABI4-MtEDN3
MtbZIP-102-MtABI5
IAA4-5-SHY2
M0766_1.02
M1363_1.02
M1656_1.02
M2371_1.02
M2354_1.02
MtRRB15
M0853_1.02
M0251_1.02
IBM1
M0758_1.02
MTF1
M0011_1.02*
M1351_1.02
M0023_1.02*
IBM1
Time
Module 5
M0012_1.02*
M0252_1.02
M1569_1.02
M1408_1.02
M1211_1.02*
M0854_1.02
M1321_1.02
VND7-SMB
M1274_1.02
M0268_1.02
M0155_1.02
MTF1-MtMADS1
MtRRB15
M2350_1.02
M1412_1.02
IBM1
MtABI4-MtEDN3
M1701_1.02
M0044_1.02*
M0073_1.02
IBM1
M0851_1.02
M1643_1.02
BHLH
MTF1-MtMADS1
M2355_1.02
EIN3
M2353_1.02
M0582_1.02
M1564_1.02
M0076_1.02*
IBM1
M0075_1.02
MtEDN1-MtEDN2-MtEDN3
BHLH
M0036_1.02*
MtbZIP-102-MtABI5
M1410_1.02
IBM1
MtRRB15
M2631_1.02*
MTF1-MtMADS1
M2372_1.02
M1645_1.02
IAA4-5-SHY2
IBM1
ATAC
Control
15 min
30 min
1 hr
2 hr
4 hr
8 hr
24 hr
Time
Edge weight
-0.05
0
0.05
Module 6
M2381_1.02
M2662_1.02
M0025_1.02*
M1305_1.02
M1274_1.02
M0024_1.02*
MtCBF4-MtCBF3b
MtbZIP-102-MtABI5
M1569_1.02
M0252_1.02
M0854_1.02
EIN3
M0035_1.02*
M1412_1.02
IBM1
BHLH
M0012_1.02*
M0157_1.02
M1320_1.02
IBM1
HDG2
IAA4-5-SHY2
M0769_1.02
MtbZIP-102-MtABI5
M1659_1.02
MtHB1
M1566_1.02
EIN3
M0582_1.02
M0016_1.02*
M0271_1.02
M0851_1.02
MTF1-MtMADS1
M1565_1.02
M0036_1.02*
M0262_1.02
M0028_1.02*
IBM1
M0758_1.02
MTF1-MtMADS1
E2Fe-DEL1
M1559_1.02
M1405_1.02
M1651_1.02
M1572_1.02
M0036_1.02*
M0765_1.02
M1307_1.02
M2371_1.02
M1644_1.02
MTF1
M0161_1.02
IBM1
BHLH
M1310_1.02
M1932_1.02
IAA4-5-SHY2
BHLH
M1645_1.02
ATAC
IBM1
Time
Module 7
IBM1
BHLH
M2631_1.02*
M0765_1.02
M0845_1.02
M0075_1.02
M1877_1.02
MtABI3
M0762_1.02
M1564_1.02
CCAATBOX
M1337_1.02
M1569_1.02
M2372_1.02
BHLH
IBM1
M0023_1.02*
M2355_1.02
M1319_1.02
M0028_1.02*
M1405_1.02
M1407_1.02
M0005_1.02*
M0015_1.02*
M0162_1.02
M0758_1.02
IBM1
M0126_1.02*
IAA4-5-SHY2
IBM1
M1687_1.02
EIN3
M0851_1.02
IBM1
IBM1
M0251_1.02
MtHB1
M0022_1.02*
M1340_1.02
M1656_1.02
M2381_1.02
M0166_1.02
MtNAC969
M2379_1.02
M1321_1.02
M0118_1.02
M0004_1.02*
M0270_1.02
M1415_1.02
M1309_1.02
MtABI4
M0010_1.02*
M2354_1.02
M1649_1.02
M1949_1.02
MtbZIP-627
BHLH
M0263_1.02
M2662_1.02
IAA4-5-SHY2
M1365_1.02
IBM1
M1685_1.02
M1275_1.02
MtHB1
M1650_1.02
M0164_1.02
M0854_1.02
IBM1
IBM1
MtRRB15
Time
*Summary of multiple mappings of motifs to common TF names
M0027_1.02, M0004_1.02, M0005_1.02, M0010_1.02, M0022_1.02, M0025_1.02, M0024_1.02,
M0028_1.02, M2360_1.02, M0076_1.02, M0012_1.02 and M0036_1.02
-> MtERF1, MtABI4, MtEDN1, MtEDN2 and MtEDN3
M1211_1.02 -> MtKNOX8, MtKNOX1, MtKNOX7 and MtKNOX6
M0011_1.02 -> MtERF1, MtEDN1, MtEDN2 and MtEDN3
M0023_1.02 -> MtEDN1, MtEDN2 and MtEDN3
M0044_1.02 and M0016_1.02 -> MtABI4, MtEDN1, MtEDN2, MtEDN3
M2631_1.02 -> MtPLT1, MtPLT2, MtPLT5, MtPLT3, MtPLT4
M0035_1.02 -> MtERF1, MtABI4, MtEDN1 and MtEDN2
M0015_1.02 -> MtEFD2-MtEFD-MtABI4
M0126_1.02 -> MtABI4-MtEDN1-MtEDN2-MtEDN3
B
Modules
0
1
2
3
4
5
6
IBM1
MtABI4
MtEDN1
MtEDN2
MtEDN3
MtERF1
BHLH
EIN3
IAA4_5
M1569_1.02
SHY2
M0252_1.02
M0758_1.02
M0851_1.02
M0854_1.02
M1412_1.02
M1559_1.02
M1645_1.02
M2355_1.02
MtABI5
MtbZIP_102
MTF1
MtHB1
MtRRB15
ATAC
M0075_1.02
M0164_1.02
M0251_1.02
M0582_1.02
M0765_1.02
M1274_1.02
M1321_1.02
M1365_1.02
M1406_1.02
M1408_1.02
M1563_1.02
M1564_1.02
M1644_1.02
M1656_1.02
M1877_1.02
M2354_1.02
M2371_1.02
M2372_1.02
M2381_1.02
M2662_1.02
M2863_1.02
MtCBF3b
MtCBF4
MtKNOX1
MtKNOX6
MtKNOX7
MtKNOX8
MtMADS1
MtPLT1
MtPLT2
MtPLT3
MtPLT4
MtPLT5
SMB
VND7

294

Fig. 6 Regulator prioritization results. (A) Summary of DRMN edge weights across the time course that meet a *t* test criterion of significant change (P <0.01) between the first and second half of the time course. **(B)** Regulators prioritized based on the frequency with which they are selected across modules with the t-test criteria, including IBM1, EDN3, MTF1, and EIN3.

According to this criterion chromatin accessibility of gene promoters was an important predictor of expression for highly expressed genes (modules 5 and 6, **Fig. 6A**). We also identified the TFs IBM1 (Increase in BONSAI methylation 1), ERF1 (ETHYLENE RESPONSE FACTOR 1), EDN1-3 (ERF differentially regulated during nodulation 1, 2 and 3), EIN3 (Ethylene Insensitive 3), SHY2 (SHORT HYPOCOTYL 2), ABI4-5 (ABSCISIC ACID-INSENSITIVE 4 and 5), MTF1 (MAD-box Transcription Factor 1), MtRRB15 (type-B response regulator 15), as well as several markers of meristem cells of KNOX and PLT (PLETHORA) protein families (**Fig. 6B**), as important regulators; all of which are involved in nodulation. We additionally tested the inferred modules for enrichment of Gene Ontology (GO) processes (**Fig. S5**). Several processed relevant to N-fixing symbiosis were identified, including nodule morphogenesis, root-hair elongation and the MAPK cascade, as well as others relating to regulation and chromatin organization.

Identification of the targets of DRMN-prioritized regulators

DRMN identified regulators of gene expression dynamics in response to LCOs. Next, we aimed to identify their targets using two criteria based on (1) DRMN modules and motif enrichment, and (2) DRMN transitioning gene sets and regulatory motifs selected by a regularized regression method, Multi-Task Group LASSO (MTG-LASSO).

Targets identified based on DRMN modules and motif enrichment.

The first criteria to define a given regulator's targets was based on identifying genes belonging to a DRMN module controlled by that regulator, based on the DRMN regression models. Only genes containing the motif to which the regulator binds to in their promoter were considered targets.

Moreover, only genes that belong to a module enriched for the associated motif were considered (q-value <0.05). Using these criteria we predicted targets of DRMN-prioritized regulators IBM1, ERF1, EDN1-3, EIN3, SHY2, ABI4-5, MTF1 and KNOX and PLT family members. We also determined putative targets of known regulatory genes such as *NIN*, *ERN1*, or *NF-YA1*.

Targets identified based on DRMN transitioning gene sets and MTG-LASSO

Expression-based network inference is commonly used to define regulator-gene relationships (32), but was challenging here due to the small number of samples. Therefore, we applied a computational method with a structured sparsity approach, Multi-Task Group LASSO (MTG-LASSO) (SLEP v4.1 package (33), **Fig. 7A**) to identify regulators (motifs/TFs) for each of the 79 DRMN transitioning gene sets (10,176 genes). Here the same feature data from the DRMN analysis was used. This approach was selected because it was more amenable to our sparse dataset than per-gene regression (34, 35).

We determined appropriate MTG-LASSO parameter settings for each transitioning gene set and established criteria for regulator-target predictions. 33 (of 79) transitioning gene clusters were significantly associated with regulatory motifs with MTG-LASSO, generating 122,245 regulatory edges connecting 126 regulatory motifs to 5,978 genes (**Fig. S7**).

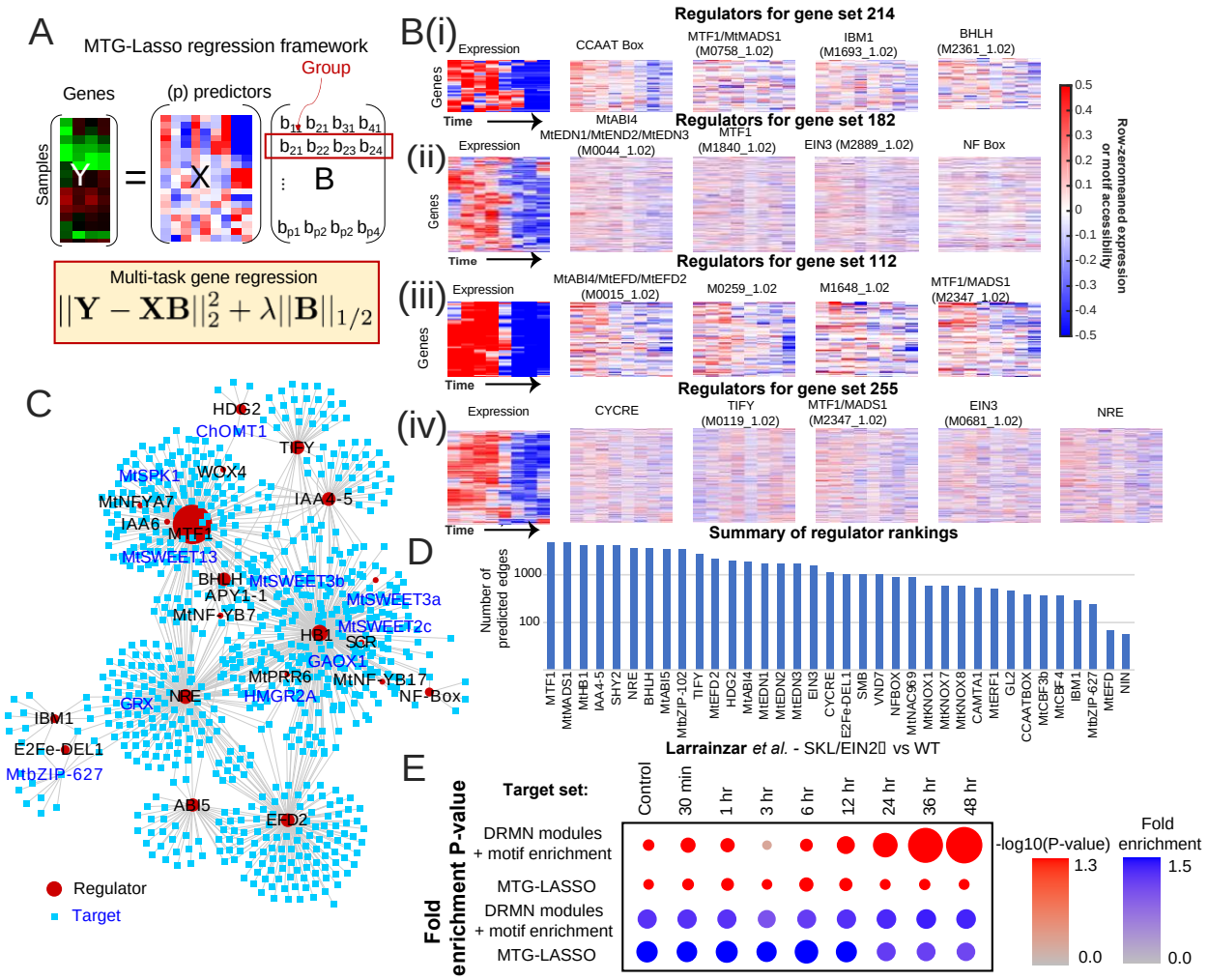


Fig. 7 Multi-task group (MTG) LASSO to predict regulators of transitioning genes. (A) MTG Lasso was applied to infer significant regulatory features for each transitioning gene set. Shown is a model of predicting Y (expression) for a set of genes using the features X for the genes and coefficients B. MTG-LASSO picks regulators for a gene set. **(B)** Multiple gene sets have notable trends for regulators of interest from the module-level analysis (IBM1, MTF1, EIN3, EDN3). Shown are four gene sets presenting patterns of association between gene expression and motif accessibility profiles. **(C)** Visualization of the top 1000 predicted TF gene network edges. **(D)** Ranking of regulators by the number of predicted targets highlights both regulators identified at the module level, and those that are not such as TIFY and CYCRE. **(E)** Overlap statistics of predicted EIN3 targets with differentially expressed genes from the Larrainzar et al. SKL/EIN2 mutant and wild-type time-course data set. Overlap is quantified with the -log10(p) of hypergeometric test *P* values (red color scale, and increasing size reflects increasing significance beyond p<0.05 or -log10(p)>1.3), and fold-enrichment (below, blue color scale, and size scale with the magnitude of the fold-enrichment score).

Several gene sets exhibit consistent downregulation of expression and corresponding reduction in accessibility of predicted regulatory motifs (**Fig. 7B**). For example, gene set 214 (57 genes) shows downregulation of gene expression and reduced motif accessibility (after 4 hr) for several TFs ranking among those with the greatest number of edges: IBM1, MTF1 and BHLH (**Fig. 7, B and D**). Similarly, gene set 182 was predicted to be regulated by EDN3, MTF1, EIN3, and NF-Box motif, and exhibited correlated trends between gene expression and regulatory feature accessibility (**Fig. 7B**).

Validation of predicted targets by comparison with published data

To validate our predicted targets, we tested the overlap between predicted targets based on (1) DRMN module-motif enrichment, and (2) MTG-LASSO application to DRMN transitioning gene sets, to the differentially expressed genes from the Larrainzer, et al. experiment comparing wild-type and *skl/ein2* mutant strains of *M. truncatula*(12). We selected the *skl/ein2* mutant data to test the targets of EIN3 because EIN2 and EIN3 function are closely related (36), and EIN3 was implicated as a significant regulator in the DRMN analysis. We find substantial overlap (hypergeometric test $P < 0.05$) between the differentially expressed genes identified in mutant/wild-type comparisons at all time points, and our predicted targets (**Fig. 7E**). The significant overlap between DRMN/MTG-LASSO-predicted targets of EIN3 and the experimentally measured targets of EIN2 supports the validity of the predicted targets, despite differing experimental techniques.

DISCUSSION

The enormous economic and environmental cost of plant nitrogen fertilization motivates efforts towards identifying molecular mechanisms underlying legume perception of nitrogen-fixing

bacteria, and nodule development. Understanding such mechanisms establishes the foundation for introducing this process in non-nitrogen-fixing species. We dissected the gene regulatory network in *M. truncatula* roots in response to *S. meliloti* LCOs by jointly profiling the temporal changes in the transcriptome and chromatin accessibility, and integrating these data computationally. Extensive changes in the transcriptome have been known to occur in *Medicago* roots in response to rhizobia signals and we show these changes are accompanied and facilitated by extensive chromatin remodeling. While the overall percentage of accessible chromatin regions remained similar across our time course experiment, regions of accessibility underwent a dramatic shift, particularly 1-4 hr after treatment. This remodeling appears to anticipate the development of root nodules, which requires stringent temporal and spatial control of gene expression. Chromatin accessibility of gene promoters notably also emerged as a significant predictor of gene expression (**Fig. 6**). These changes in chromatin accessibility enable and enhance the transcriptional changes required for nodule development by providing promoter access to regulators that may be inactive in other stages of plant development. Correlation was additionally observed between gene expression and promoter chromatin accessibility profiles of several essential regulators of nodulation, including ERN1, CRE1, SKL/EIN2, IDP3/CYCLOPS, and ERN2. Close coordination between chromatin accessibility and gene expression in LCO response is likely essential for root nodule development.

We applied novel methods for time-series analysis, ESCAROLE and DRMN (37), to model temporal changes in gene expression and regulatory motif accessibility. ESCAROLE enabled us to characterize the transcriptional dynamics, while DRMN allowed us to predict which TFs are most important for predicting expression dynamics. In tune with the theme of chromatin reorganization under LCO treatment response, DRMN identified IBM1 as a critical regulator.

IBM1 encodes a JmjC domain-containing histone demethylase that catalyzes the removal of H3K9 methylation and di-methylation in Arabidopsis (38). DRMN also identified regulatory genes involved in hormone responses associated with the early steps of symbiosis and nodule formation such as ethylene (ERF1, EDN1-3 and EIN3) and ABA (ABI4-5). EIN3 is a transcription factor mediating ethylene-regulated gene expression and morphological responses in *Arabidopsis*. The role of EIN3 in rhizobium-legume symbiosis or LCOs signaling remains uncharacterized, but *sickle* (*skl*) mutants for an EIN2 ortholog develop more infection threads and nodules and respond more to LCOs than wild-type plants, and ethylene treatment inhibits LCO signaling and nodule formation (39). ABI4 and ABI5, basic leucine zipper transcription factors implicated in several plant functions, coordinate LCO and cytokinin signaling during nodulation in *M. truncatula* (40). DRMN also identified regulators associated with the hormones involved in the nodule initiation, auxin (SHY2) and cytokinin (MtRRB15). SHY2, a member of the Aux/IAA family, plays a critical role in cell differentiation at root apical meristem and is activated by cytokinin (41, 42). SHY2 was proposed as a candidate for nodule meristem regulation and differentiation after showing a very localized expression pattern in the nodule meristematic region (43). Also related to nodule meristem initiation, KNOX TF-family members and PLT1-5 were predicted as regulators of gene expression in response to LCOs. *MtPLT* genes (*MtPLT1-5*) are part of the root developmental program recruited from root formation, and control meristem formation and maintenance for root and nodule organogenesis (44).

Finally, our analysis predicted genome-wide targets for transcription factors, including novel regulators identified by DRMN and previously characterized essential regulators of root nodulation, such as NIN, NF-YA1/NF-YB1, and CYCLOPS. This data established specific hypotheses about the regulation of gene expression in LCO response, informing future

experimental validation of regulator-target relationships. For example, among novel regulators identified by DRMN, MTG-Lasso analysis pointed to *NIN* as a direct target of SHY2 and MTF1, and *FLOT4*, required for infection thread formation, as a target of IBM1 (45). Among already known regulators, MTG-Lasso indicated that *ARF16a* and *SPK1* are targets of NF-Y TFs. ARF16a and SPK1 control infection initiation and nodule formation (1). Several NF-Y genes (NF-YA5 and NF-YB17) were identified as regulated by CYCLOPS. In summary, our novel transcriptomic and accessibility datasets and computational framework to integrate these datasets provide a valuable resource to the plant community to understand the gene regulatory programs controlling the establishment of root nodulation symbiosis in *Medicago truncatula*.

MATERIALS AND METHODS

Detailed information, including plant material, growth conditions and treatments, ATAC-seq and RNA-seq library preparation and sequencing, differential expression analysis, ESCAROLE clustering, exploratory analysis of chromatin accessibility, Dynamic Regulatory Module Networks, and target prediction and validation analyses, are provided in Supplementary Information (SI) appendix (Materials and Methods). All data are described/linked in the SI Appendix.

ACKNOWLEDGEMENTS

Funding: This work was supported by the Department of Energy Office of Science Biological and Environmental Research (DE-SC0018247 to M.K. and J.-M.A.). We acknowledge the support of the Center of High Throughput Computing at UW-Madison to enable computational experiments. **Author contributions:** S.R., J.-M.A., M.K., S.A.K. and D.C. conceived and

designed the study. D.C., K.M.B., T.B.I., L.G.M., P.M.T., C.D., S.C. and H.W.S. planned and carried out the wet laboratory experiments. J.M. developed biological resources (LCO). S.A.K., S.R. and W.J.P. carried out bioinformatic analysis. S.R., J.-M.A., M.K., S.A.K. and D.C. drafted the manuscript. All coauthors discussed and interpreted the results and contributed to the final version of the manuscript. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** The sequencing data presented in this publication have been deposited in NCBI's Gene Expression Omnibus and are accessible through GEO Series accession number GSE154845 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE154845). The results from DRMN and MTG-LASSO gene target predictions are available in Zenodo (https://www.zenodo.org/deposit/4057115).

REFERENCES

1. S. Roy, *et al.*, Celebrating 20 Years of Genetic Discoveries in Legume Nodulation and Symbiotic Nitrogen Fixation. *Plant Cell* **32**, 15–41 (2020).

2. V. Smil, Nitrogen in crop production: An account of global flows. *Glob. Biogeochem. Cycles* (1999) https:/doi.org/10.1029/1999GB900015.

3. A. Genre, G. Russo, Does a common pathway transduce symbiotic signals in plant–microbe interactions? *Front. Plant Sci.* (2016) https:/doi.org/10.3389/fpls.2016.00096.

4. E. Limpens, *et al.*, LysM Domain Receptor Kinases Regulating Rhizobial Nod Factor-Induced Infection. *Science* (2003) https:/doi.org/10.1126/science.1090074.

5. J. F. Arrighi, *et al.*, The Medicago truncatula lysine motif-receptor-like kinase gene family includes NFP and new nodule-expressed genes. *Plant Physiol.* (2006) https:/doi.org/10.1104/pp.106.084657.

6. G. E. D. Oldroyd, Speak, friend, and enter: Signalling systems that promote beneficial symbiotic associations in plants. *Nat. Rev. Microbiol.* (2013) https:/doi.org/10.1038/nrmicro2990.

7. S. Kim, *et al.*, Ca2+-regulated Ca2+ channels with an RCK gating ring control plant symbiotic associations. *Nat. Commun.* **10**, 3703 (2019).

8. W. Capoen, *et al.*, Nuclear membranes control symbiotic calcium signaling of legumes. *Proc. Natl. Acad. Sci.* **108**, 14348–14353 (2011).

9. J. Levy, A Putative Ca2+ and Calmodulin-Dependent Protein Kinase Required for Bacterial and Fungal Symbioses. *Science* **303**, 1361–1364 (2004).

10. G. E. D. Oldroyd, J. D. Murray, P. S. Poole, J. A. Downie, The Rules of Engagement in the Legume-Rhizobial Symbiosis. *Annu. Rev. Genet.* (2011) https:/doi.org/10.1146/annurev-genet-110410-132549.

11. M. Venkateshwaran, J. D. Volkening, M. R. Sussman, J. M. Ané, Symbiosis and the social network of higher plants. *Curr. Opin. Plant Biol.* (2013) https:/doi.org/10.1016/j.pbi.2012.11.007.

12. E. Larrainzar, *et al.*, Deep sequencing of the Medicago truncatula root transcriptome reveals a massive and early interaction between nodulation factor and ethylene signals. *Plant Physiol.* (2015) https:/doi.org/10.1104/pp.15.00350.

13. C. Satgé, *et al.*, Reprogramming of DNA methylation is critical for nodule development in Medicago truncatula. *Nat. Plants* (2016) https:/doi.org/10.1038/nplants.2016.166.

14. M. G. Poirier, M. Bussiek, J. Langowski, J. Widom, Spontaneous Access to DNA Target Sites in Folded Chromatin Fibers. *J. Mol. Biol.* (2008) https:/doi.org/10.1016/j.jmb.2008.04.025.

15. M. F. Wu, *et al.*, SWI2/SNF2 chromatin remodeling ATPases overcome polycomb repression and control floral organ identity with the LEAFY and SEPALLATA3 transcription factors. *Proc. Natl. Acad. Sci. U. S. A.* (2012) https:/doi.org/10.1073/pnas.1113409109.

16. B. Sun, *et al.*, Timing mechanism dependent on cell division is invoked by Polycomb eviction in plant stem cells. *Science* (2014) https:/doi.org/10.1126/science.1248559.

17. M. Nagymihály, *et al.*, Ploidy-dependent changes in the epigenome of symbiotic cells correlate with specific patterns of gene expression. *Proc. Natl. Acad. Sci. U. S. A.* (2017) https:/doi.org/10.1073/pnas.1704211114.

18. L. F. Czaja, *et al.*, Transcriptional responses toward diffusible signals from symbiotic microbes reveal MtNFP- and MtDMI3-dependent reprogramming of host gene expression by arbuscular mycorrhizal fungal lipochitooligosaccharides. *Plant Physiol.* (2012) https:/doi.org/10.1104/pp.112.195990.

19. C. M. Rose, *et al.*, Rapid phosphoproteomic and transcriptomic changes in the rhizobia-legume symbiosis. *Mol. Cell. Proteomics* (2012) https:/doi.org/10.1074/mcp.M112.019208.

20. D. Chasman, *et al.*, Inferring Regulatory Programs Governing Region Specificity of Neuroepithelial Stem Cells during Early Hindbrain and Spinal Cord Development. *Cell Syst.* **9**, 167-186.e12 (2019).

21. C. Wasternack, B. Hause, Jasmonates: biosynthesis, perception, signal transduction and action in plant stress response, growth and development. An update to the 2007 review in Annals of Botany. *Ann. Bot.* **111**, 1021–1058 (2013).

22. L.-Q. Chen, *et al.*, *ATX3* , *ATX4* , and *ATX5* Encode Putative H3K4 Methyltransferases and Are Critical for Plant Development. *Plant Physiol.* **174**, 1795–1806 (2017).

23. C. Xu, F. Luo, F. Hochholdinger, LOB Domain Proteins: Beyond Lateral Organ Boundaries. *Trends Plant Sci.* **21**, 159–167 (2016).

24. T. Soyano, Y. Shimoda, M. Kawaguchi, M. Hayashi, A shared gene drives lateral root development and root nodule symbiosis pathways in *Lotus*. *Science* **366**, 1021–1023 (2019).

25. O. Valdés-López, *et al.*, A Novel Positive Regulator of the Early Stages of Root Nodule Symbiosis Identified by Phosphoproteomics. *Plant Cell Physiol.* **60**, 575–586 (2019).

26. S. Gonzalez-Rizzo, M. Crespi, F. Frugier, The *Medicago truncatula* CRE1 Cytokinin Receptor Regulates Lateral Root Development and Early Symbiotic Interaction with *Sinorhizobium meliloti*. *Plant Cell* **18**, 2680–2693 (2006).

27. Y. Zhang, *et al.*, Model-based Analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).

28. A. Fotuhi Siahpirani, *et al.*, "Dynamic regulatory module networks for inference of cell type specific transcriptional networks" (Bioinformatics, 2020) https:/doi.org/10.1101/2020.07.18.210328 (July 20, 2020).

29. T. G. do Rego, H. G. Roider, F. A. T. de Carvalho, I. G. Costa, Inferring epigenetic and transcriptional regulation during blood cell development with a mixture of sparse linear models. *Bioinformatics* **28**, 2297–2303 (2012).

30. R. Caruana, Multitask Learning. *Mach. Learn.* **28**, 41–75 (1997).

31. M. T. Weirauch, *et al.*, Determination and Inference of Eukaryotic Transcription Factor Sequence Specificity. *Cell* **158**, 1431–1443 (2014).

32. S. Roy, *et al.*, Integrated Module and Gene-Specific Regulatory Inference Implicates Upstream Signaling Networks. *PLoS Comput. Biol.* **9**, e1003252 (2013).

33. J. Ye, J. Liu, S. Ji, SLEP: Sparse Learning with Efficient Projections. *Ariz. State Univ.* (2010).

34. V. A. Huynh-Thu, A. Irrthum, L. Wehenkel, P. Geurts, Inferring Regulatory Networks from Expression Data Using Tree-Based Methods. *PLoS ONE* **5**, e12776 (2010).

35. A.-C. Haury, F. Mordelet, P. Vera-Licona, J.-P. Vert, TIGRESS: Trustful Inference of Gene REgulation using Stability Selection. *BMC Syst. Biol.* **6**, 145 (2012).

36. F. An, *et al.*, Ethylene-Induced Stabilization of ETHYLENE INSENSITIVE3 and EIN3-LIKE1 Is Mediated by Proteasomal Degradation of EIN3 Binding F-Box 1 and 2 That Requires EIN2 in *Arabidopsis*. *Plant Cell* **22**, 2384–2401 (2010).

37. A. Fotuhi Siahpirani, *et al.*, "Dynamic regulatory module networks for inference of cell type specific transcriptional networks" (Bioinformatics, 2020) https:/doi.org/10.1101/2020.07.18.210328 (July 20, 2020).

38. S. Inagaki, *et al.*, Autocatalytic differentiation of epigenetic modifications within the Arabidopsis genome. *EMBO J.* **29**, 3496–3506 (2010).

39. G. E. D. Oldroyd, E. M. Engstrom, S. R. Long, Ethylene Inhibits the Nod Factor Signal Transduction Pathway of *Medicago truncatula*. *Plant Cell* **13**, 1835–1849 (2001).

40. Y. Ding, *et al.*, Abscisic Acid Coordinates Nod Factor and Cytokinin Signaling during the Regulation of Nodulation in *Medicago truncatula*. *Plant Cell* **20**, 2681–2695 (2008).

41. F. Ariel, *et al.*, Two Direct Targets of Cytokinin Signaling Regulate Symbiotic Nodulation in *Medicago truncatula*. *Plant Cell* **24**, 3838–3852 (2012).

42. H. Takatsuka, M. Umeda, Hormonal control of cell division and elongation along differentiation trajectories in roots. *J. Exp. Bot.* **65**, 2633–2643 (2014).

43. B. Roux, *et al.*, An integrated analysis of plant and bacterial gene expression in symbiotic root nodules using laser-capture microdissection coupled to RNA sequencing. *Plant J.* **77**, 817–837 (2014).

44. H. J. Franssen, *et al.*, Root developmental programs shape the Medicago truncatula nodule meristem. *Development* **142**, 2941–2950 (2015).

45. C. H. Haney, S. R. Long, Plant flotillins are required for infection by nitrogen-fixing bacteria. *Proc. Natl. Acad. Sci.* **107**, 478–483 (2010).