Gene Regulatory Network of Secondary Cell Wall Biosynthesis during VND7 Induced *de novo* Xylem Formation

Haonan Tong[1†], Hao Chen[2†], Cranos M. Williams[1*]

[1] Department of Electrical and Computer Engineering, North Carolina State University, Raleigh, USA.
[2] Department of Plant and Microbial Biology, North Carolina State University, Raleigh, USA.

* Corresponding author. email: cmwilli5@ncsu.edu

† These authors contributed equally to this work.
Manuscript submitted June 11, 2021; accepted August 24, 2021.
doi: 10.17706/ijbbb.2021.11.4.74-81

Abstract: Accumulation of -omics data allows us to analyze the coordination and cooperation of multiple genes involved in different biological processes. Gene regulatory networks (GRNs) are used to characterize the regulatory relationships between transcription factors and downstream genes involved in different biological processes. The secondary cell wall (SCW) formation is involved with many important biological processes in plants, such as stress defense, mechanical reinforcement, and the transportation of water and nutrients. We construct GRNs based on the time-series data of VND7-induced *de novo* SCW formation using multiple algorithms, and then evaluate each GRN model based on the MYB46 experimental validated regulation data. From the optimal GRN model, we not only identify 8 TFs that have been previously demonstrated as the master regulators of SCW formation, but also show 6 novel SCW regulators which include EGL3, DREB19, TCP14, BZIP61, RGA2, and a zinc-finger type TF. From further *in silico* annotation of the downstream genes that are regulated by these TFs, we find the shared transcriptional program between the SCW formation and the processes of photosynthesis and drought response. Overall, our work suggests a pipeline for reconstructing and analyzing GRN to pinpoint gene functions in biological processes.

Key words: Gene regulatory network, secondary cell wall, gene expression, *Arabidopsis*.

1. Introduction

Vascular cell differentiation and secondary cell wall (SCW) biosynthesis are closely related biological processes because both processes can be regulated by the same signaling pathway [1], [2]. Many transcription factors (TFs) regulate both vascular cell differentiation and cell-wall component biosynthesis. Several NAC domain TFs have been identified to not only specify xylem tissue identity by regulating vessel and fiber cell differentiation, but also directly and indirectly regulate secondary cell wall component genes. These TFs include NST1 (NAC SECONDARY WALL THICKENING PROMOTING FACTOR1), NST3/SND1 (SECONDARY WALL-ASSOCIATED NAC DOMAIN PROTEIN 1), and VND6/7 (VASCULAR-RELATED NAC-DOMAIN6/7). In *Arabidopsis*, the NAC domain TFs VND6 and VND7 control secondary wall formation and programmed cell death (PCD) of the vessels in both root and shoot tissues [3], [4]. The overexpression of VND7, along with the activation of many secondary wall genes, [5] induces the differentiation of *Arabidopsis* seedling cells without SCW into protoxylem with SCW. The VND7-induced xylem differentiation system in *Arabidopsis* seedlings provides us with an opportunity for elucidating the gene regulatory

network (GRN) underlying the *de novo* SCW formation.

Computational algorithms that decipher GRNs from expression data [6], [7] could help us better understand the transcriptional program regulating SCW formation. The GRN models can be constructed by correlation-based and machine learning-based methods [8]. Correlation-based methods detect either linear (Pearson rank coefficient) or nonlinear (Spearman rank coefficient) monotonic relationships from gene expression [9] and are well suited for global network construction from RNA-seq data [10]. Recently, machine learning-based GRN algorithms including tree-based methods (GENIE3 [11], dynGENIE3 [12]) are getting popular. While studies suggest supervised learning strategies for inferring genome-scale networks remain elusive [13], they have gained attention for their high prediction accuracy in *in silico* network studies [14] and successful application in gene expression analysis [15], [16].

In this study, we established and optimized a GRN pipeline for understanding transcriptional regulatory mechanisms for SCW formation (Fig. 1). We reconstructed GRNs for SCW formation from 9 transcriptomic time-points (0, 1, 3, 6, 9, 12, 24, 30, and 48 h) of *Arabidopsis* seedlings after the induction of VND7.

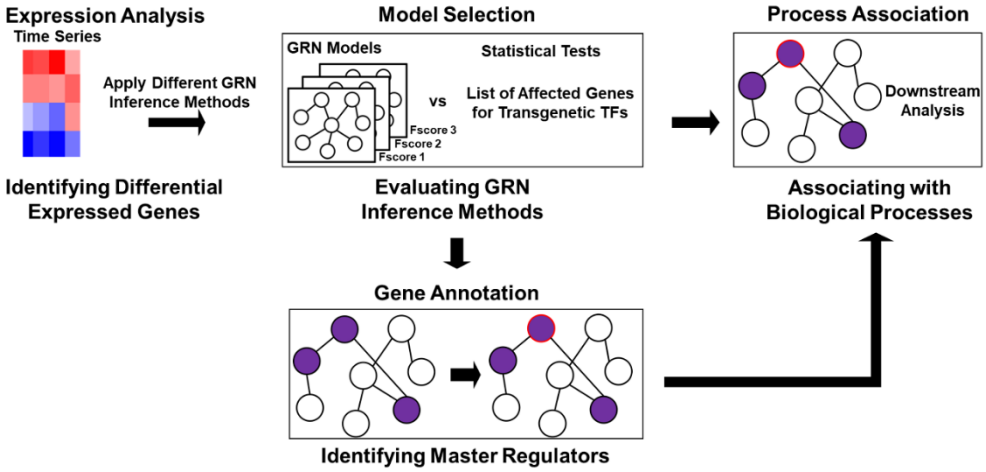


Fig. 1. Workflow of GRN construction and analyses.

Overexpression using two correlation metrics (Pearson rank coefficient, Spearman rank correlation) and two machine learning tools (GENIE3 [11], dynGENIE3 [12]). MYB46 experimental data was used to evaluate these networks. We found that GENIE3-based network attained the highest F-score with appropriate network size. By further analyzing the GENIE3 SCW-related GRN, we identified six novel SCW regulators as EGL3, DREB19, TCP14, BZIP61, RGA2, and a zinc-finger type TF. Biological processes enriched in the downstream genes of the six TFs that are co-regulated with SCW formation are identified from *in silico* functional annotation.

2. Materials and Methods

2.1. Microarray Data Analysis

Microarray data from the Affymetrix ATH1 array in three biological replicates at nine time points, 0, 1, 3, 6, 12, 24, 30, 48h were downloaded from Gene Expression Omnibus under accession number GSE77153, and the microarray data processing and normalization were done as in [17], [18].

2.2. Differential Analysis

Differential expression was identified between consecutive time points by a linear model [19], with fold change > 2 or <-2 and a significance level smaller than 0.05 (FDR corrected).

2.3. Network Construction

The absolute correlation coefficients and connectivity scores were calculated for ranking TF-Gene relationships. The Pearson rank coefficient and Spearman rank coefficient was used to measure correlation coefficients. GENIE3 [11] and dynGENIE3 [12] with default parameters were used to calculate connectivity scores.

2.4. Model Selection

The framework from Olsen *et al.* was used to select the final GRN model [20]. Using the reference's proposed statistical analyses, 599 genes that are affected by MYB46 were collected (283 genes are affected by MYB46 overexpression, 187 genes affected by MYB46 knockout, and 184 direct targets identified using the estradiol-activated MYB46 in the presence of cycloheximide) [21]-[23]. Downstream genes in the MYB46 were extracted from each GRN model and were classified as true positive (TP), false positive (FP) and false negative (FN). The quality measure F-score is calculated for a GRN model

$$F = \frac{2TP}{2TP + FP + FN}, F \in [0,1]$$

where F=1 corresponds to no misclassification and F=0 to no true detections.

2.5. GO Enrichment Analysis

GO terms were annotated and depicted for each gene using AgriGO2.0 [24]. AgriGO2.0 then was used to perform GO enrichment. The GO terms with an FDR < 0.05 were selected for further analysis, and the redundant terms were filtered out.

3. Results and Discussion

3.1. GRN Model Inferred from GENIE3 Outperforms Other Selected Algorithms in Predicting MYB46 Downstream Genes

The cells in seedlings of the inducible VND7 lines have been showed to generate *de novo* secondary cell wall within 48h after DEX-induced VND7 overexpression [17]. The transcriptomic data of seedlings collected at nine time points within these 48 h (0, 1, 3, 6, 9, 12, 24, 30, and 48 h) were used to identify differential expressed genes (DEGs). Finally, we extracted a subset of 4,214 DEGs (FDR < 0.05, fold change >2 or <-2), which includes 413 TF genes. We measured the connectivity strength of transcriptional relationships between each of 413 differential expressed TFs and their regulated genes using Pearson rank coefficient, Spearman rank coefficient, GENIE3 and dyGENIE3. As a result, we ranked each of 1,739,969 TF-gene pairs with a score based on each of four methods. For each method, we selected a subset of TF-gene pairs using thresholds at the 75th, 90th and 95th percentile of the scores, that is, the top 25%, 10%, 5% gene connections. Totally, we generated 12 GRN models from 4 different methods using 3 percentile thresholds. To evaluate which GRN model fits the experimental data best, we collected 599 experimentally validated genes whose expression is affected by MYB46 [21]-[23]. Based on this experimental data, we analyzed F-scores of MYB46 subnetworks from 12 networks (Table 1). We found that GRN based Spearman rank coefficient has the best F-score at both 75th and 90th percentiles, followed by networks based on Pearson rank coefficient and GENIE3. At the 95th percentile, GENIE3 has the highest F-score compared to other methods. dynGENIE3 performs the worst across different percentile cut-offs, suggesting that the ODE-based model would likely not be competitive in high dimensional GRN inference. We also found that networks with more connections (network with low percentile thresholds) could lead to higher F-score. However, high number of connections create difficulty in visualizing the network. To balance the network

size and true detection rate, we used the GENIE3-based GRN with top 5% gene connections for identifying potential master regulators in SCW formation.

Table 1. F-Score Measurement for Model Selection

Methods	75th Percentile	90th Percentile	95th Percentile
Pearson Rank Coefficient	0.0932	0.0818	0.0568
Spearman Rank Coefficient	0.0932	0.0860	0.0672
GENIE3	0.0843	0.0808	0.0826
dynGENIE3	0.0720	0.0610	0.0256

3.2. EGL3, DREB19, TCP14, BZIP61, RGA2, and a Zinc-Finger Type TF are Identified as Potential Master TFs in SCW Formation

In the process of VND7 induced *de novo* SCW formation, 22 genes encoding the SCW biosynthetic enzymes (IRX6, GUT1, IRX9, IRX11, GUX2, CESA7, PAL4, CESA8, HCT, DUF1218, GXM1, CCOAOMT7, GLP10, PAL1, 4CL1, GUX1, LAC17, BGLU46, CESA4, CAD4, DUF547, FLY1) are differentially expressed (FDR < 0.05, fold change >2 or <-2). From the GENIE3-based GRN with top 5% gene connections, we identified 5 TFs with the highest GENIE3 ranking scores as the regulators for each of the 22 SCW genes, and these TFs served as first layer components that directly regulate 22 SCW biosynthetic enzyme genes. Then we applied the same criteria used to generate the first layer to identify regulators at the second layer that indirectly regulate 22 SCW biosynthetic enzyme genes. Finally, we constructed a multi-layered network including 22 SCW enzyme genes in the bottom layer, 217 TFs found in the next two levels, and 488 regulatory connections (edges) between regulators and target genes (Fig. 2). Given that the observable SCW associated metabolites start to be detected at 9 to12h [17], we focused on the TFs that constitutively changed at earlier time points (TFs that are included in the set of 1h/0h, 3h/1h, and 6h/3h DEGs). Finally, we identified 17 TFs, and 14 of these 17 TFs appeared in our two-layered network. Consistent with previous studies [25]-[27], MYB46/83, LBD15/18, ASL19, SND3, C3H14, and BLH10 directly and indirectly regulate SCW biosynthetic enzyme genes. From the network, we also identified TFs that have not been widely discovered before as potential SCW regulators. These novel SCW regulators included EGL3, DREB19, TCP14, BZIP61, RGA2, and a zinc-finger type TF. Our networks showed that RGA2, a repressor for gibberellin acid (GA) signaling, directly regulates cellulose synthases and hemicellulose synthesis enzymes (IRX6, GUT1, IRX9, IRX11), and indirectly regulates enzymes for lignin, cellulose and hemicellulose (GUX2, CESA7, PAL4, CESA8, HCT, DUF1218, GXM1, CCOAOMT7, GLP10, PAL1, 4CL1, GUX1, LAC17, BGLU46, CESA4, CAD4, DUF547, FLY1), which could provide a clue for elucidating the mechanism of how GA regulates SCW [28]. Another hormone-related TF, DREB19, that can be induced by JA, ABA and multiple abiotic stresses [29], indirectly regulates all of the 22 SCW biosynthetic enzymes, which suggested its potential role in regulating SCW remodeling in response to environmental stimuli. Additionally, our networks also showed that the EGL3, a key TF regulating trichome and seed coat development [30], [31], also indirectly regulates most of the 22 SCW biosynthetic enzymes. Next, we asked what other biological processes these novel SCW TFs regulate besides the SCW biosynthetic process. Identification of these biological processes co-regulated with SCW formation by the same set of TFs could help better understand how plants balance SCW formation and other growth and defense activities.

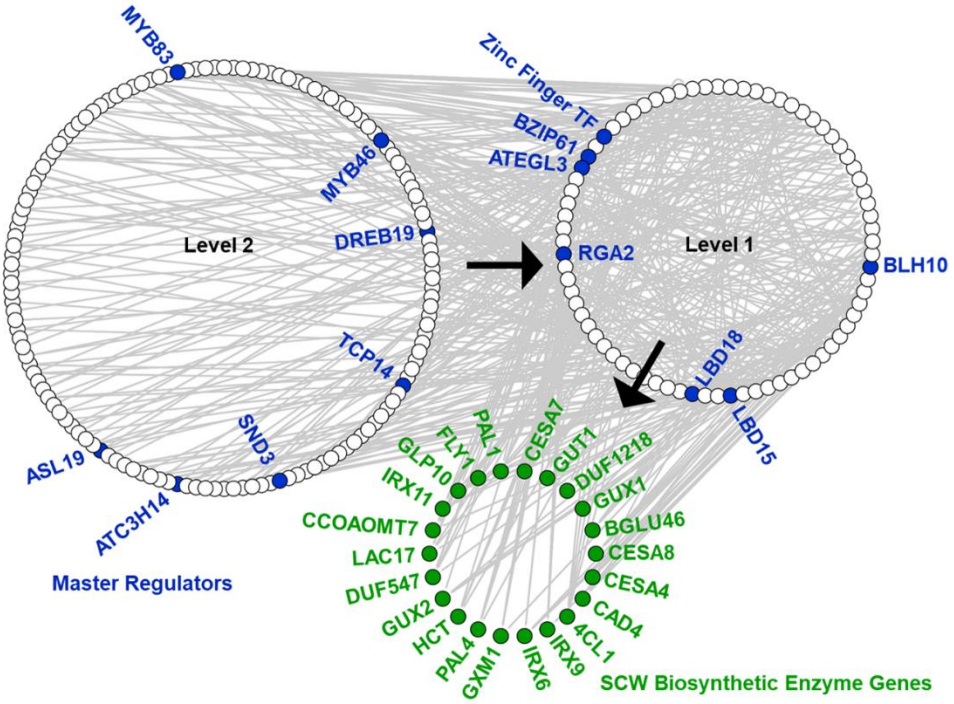


Fig. 2. GRN regulating 22 SCW genes. Blue nodes represent identified hub TFs in SCW formation. Green nodes represent SCW biosynthetic enzyme genes.

3.3. Biological Processes of Photosynthesis and Drought Response are Enriched in SCW Formation

To identify genes regulated by each of the 6 novel SCW TFs, we extracted each TF-downstream subnetwork from the GENIE3-based GRN with top 5% gene connections. The resulting downstream subnetworks of RGA2, EGL3, BZIP61, DREB19, TCP14, and zinc-finger type TF include 840 genes, 244 genes, 569 genes, 372 genes, and 794 genes, respectively. We used GO enrichment [24] to categorize the downstream genes of each TF into specific biological terms. Surprisingly, except the GO terms related to SCW formation, 4 of these 6 TFs (RGA2, DREB19, TCP14, and C2H2 zinc finger-type TF) commonly participate in the biological processes "response to light stimulus", "regulation of photosynthesis", and "carbohydrate metabolic". Given that the photosynthesis and carbohydrate metabolism create resources and energy used for SCW formation, further genetic and molecular analysis of these TFs could help understand how carbohydrate metabolism and photosynthesis affect SCW formation at a transcriptional level. Another common process regulated by 4 TFs (RGA2, TCP14, BZIP61, and C2H2 zinc finger-type TF) is "response to water deprivation". Physiological and chemical analyses of water deprivation induced SCW morphological and composition changes have been shown [32], [33]. However, the molecular and transcriptional mechanism regulated by crosstalk between SCW formation and water deprivation is not clear. Our constructed RGA2, TCP14, BZIP61, and C2H2 zinc finger-type TF directed subnetwork would be a good starting point for dissecting transcriptional regulation about the interaction of water deprivation and SCW formation.

4. Conclusion

The formation of secondary cell wall (SCW) is an important biological process. We constructed and analyzed GRNs associated with this process and provided insights into the functions of potential master TFs in SCW formation. We identified 6 novel SCW regulators including EGL3, DREB19, TCP14, BZIP61, RGA2, and a zinc-finger type TF that regulate biological processes of photosynthesis and drought response during

the SCW formation. While we have focused our efforts on understanding SCW formation, the pipeline proposed can be used for understanding other biological processes.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

Haonan Tong and Hao Chen contributed equally to the manuscript. Cranos M. Williams provided feedback, advice and comments on the manuscript.

Acknowledgment

Haonan Tong and Hao Chen were funded by the National Science Foundation (NSF IOS-1444561).

References

[1] Ito, Y., Nakanomyo, I., Motose, H., Iwamoto, K., Sawa, S., Dohmae, N., *et al.* (2006). Dodeca-CLE peptides as suppressors of plant stem cell differentiation. *Science, 313(5788)*, 842-845.

[2] Etchells, J. P., & Simon, R. T. (2010). The PXY-CLE41 receptor ligand pair defines a multifunctional pathway that controls the rate and orientation of vascular cell division. *Development, 137(5)*, 767-774.

[3] Kubo, M., Udagawa, M., Nishikubo, N., Horiguchi, G., Yamaguchi, M., Ito, J., *et al.* (2005). Transcription switches for protoxylem and metaxylem vessel formation. *Genes & Development, 19(16)*, 1855-1860.

[4] Yamaguchi, M., Kubo, M., Fukuda, H., & Demura, T. (2008). Vascular-related NAC-DOMAIN7 is involved in the differentiation of all types of xylem vessels in *Arabidopsis* roots and shoots. *The Plant Journal, 55(4)*, 652-664.

[5] Yamaguchi, M., Goué, N., Igarashi, H., Ohtani, M., Nakano, Y., Mortimer, J. C., *et al.* (2010). Vascular-related NAC-DOMAIN6 and vascular-related NAC-DOMAIN7 effectively induce transdifferentiation into xylem vessel elements under control of an induction system. *Plant Physiology 153(3)*, 906-914.

[6] Tong, H., Madison, I., Long, T. A., & Williams, C. M. (2020). Computational solutions for modeling and controlling plant response to abiotic stresses: A review with focus on iron deficiency. *Current Opinion in Plant Biology, 57*, 8-15.

[7] Van den Broeck, L., Gordon, M., Inzé, D., Williams, C. M., & Sozzani, R. (2020). Gene regulatory network inference: Connecting plant biology and mathematical modeling. *Frontiers in Genetics, 11*, 457.

[8] Haque, S., Ahmad, J. S., Clark, N. M., Williams, C. M., & Sozzani, R. (2019). Computational prediction of gene regulatory networks in plant growth and development. *Current Opinion in Plant Biology, 47*, 96-105.

[9] Santos, S. D. S., Takahashi, D. Y., Nakata, A., & Fujita, A. (2014). A comparative study of statistical methods used to identify dependencies between gene expression signals. *Briefings in Bioinformatics, 15(6)*, 906-918.

[10] Liesecke, F., Daudu, D., Bernonville, R. D. D., Besseau, S., Clastre, M., Courdavault, V., *et al.* (2018). Ranking genome-wide correlation measurements improves microarray and RNA-seq based global and targeted co-expression networks. *Scientific Reports, 8(1)*, 1-16.

[11] Irrthum, A., Wehenkel, L., & Geurts, P. (2010). Inferring regulatory networks from expression data using tree-based methods. *PloS One, 5(9)*, e12776.

[12] Geurts, P. (2018). dynGENIE3: Dynamical GENIE3 for the inference of gene networks from time series expression data. *Scientific Reports, 8(1)*, 1-12.

[13] Maetschke, S. R., Madhamshettiwar, P. B., Davis, M. J., & Ragan, M. A. (2014). Supervised, semi-supervised and unsupervised inference of gene regulatory networks. *Briefings in Bioinformatics, 15(2)*, 195-211.

[14] Marbach, D., Costello, J. C., Küffner, R., Vega, N. M., Prill, R. J., Camacho, D. M., *et al.* (2012). Wisdom of crowds for robust gene network inference. *Nature Methods, 9(8)*, 796-804.

[15] Ramírez-González, R. H., Borrill, P., Lang, D., Harrington, S. A., Brinton, J., Venturini, L., *et al.* (2018). The transcriptional landscape of polyploid wheat. *Science, 361(6403)*, eaar6089.

[16] Walley, J. W., Sartor, R. C., Shen, Z. X., Schmitz, R. J., Wu, K. J., Urich, M. A., *et al.* (2016). Integration of omic networks in a developmental atlas of maize. *Science, 353(6301)*, 814-818.

[17] Li, Z., Omranian, N., Neumetzler, L., Wang, T., Herter, T., Usadel, B., *et al.* (2016). A transcriptional and metabolic framework for secondary wall formation in *Arabidopsis*. *Plant Physiology, 172(2)*, 1334-1351.

[18] Wu, Z., Irizarry, R. A., Gentleman, R., Martinez-Murillo, F. & Spencer, F. (2004). A model-based background adjustment for oligonucleotide expression arrays. *Journal of the American Statistical Association, 99(468)*, 909-917.

[19] Smyth, G. K. (2005). Limma: Linear models for microarray data. In R. Gentleman, V. J. Carey, W. Huber, R. A. Irizarry, & S. Dudoit (Eds.), *Bioinformatics and Computational Biology Solutions Using R and Bioconductor* (pp. 397-420). New York: Springer.

[20] Olsen, C., Fleming, K., Prendergast, N., Rubio, R., Emmert-Streib, F., Bontempi, G., *et al.* (2014). Inference and validation of predictive gene networks from biomedical literature and gene expression data. *Genomics, 103(5-6)*, 329-336.

[21] Zhong, R., Richardson, E. A., & Ye, Z. H. (2007). The MYB46 transcription factor is a direct target of SND1 and regulates secondary wall biosynthesis in *Arabidopsis*. *The Plant Cell, 19(9)*, 2776-2792.

[22] Ramírez-González, R. H., Borrill, P., Lang, D., Harrington, S. A., Brinton, J., Venturini, L., *et al.* (2018). The transcriptional landscape of polyploid wheat. *Science, 361(6403)*.

[23] Zhong, R., & Ye, Z. H. (2012). MYB46 and MYB83 bind to the SMRE sites and directly activate a suite of transcription factors and secondary wall biosynthetic genes. *Plant and Cell Physiology, 53(2)*, 368-380.

[24] Tian, T., Liu, Y., Yan, H., You, Q., Yi, X., Du, Z., *et al.* (2017). agriGO v2. 0: A GO analysis toolkit for the agricultural community, 2017 update. *Nucleic Acids Research, 45(W1)*, W122-W129.

[25] Taylor-Teeples, M., Lin, L., De Lucas, M., Turco, G., Toal, T. W., Gaudinier, A., *et al.* (2015). An *Arabidopsis* gene regulatory network for secondary cell wall synthesis. *Nature, 517(7536)*, 571-575.

[26] Zhang, J., Xie, M., Tuskan, G. A., Muchero, W., & Chen, J. G. (2018). Recent advances in the transcriptional regulation of secondary cell wall biosynthesis in the woody plants. *Frontiers in Plant Science, 9*, 1535.

[27] Chen, H., Wang, J. P., Liu, H., Li, H., Lin, Y. C. J., Shi, R., *et al.* (2019). Hierarchical transcription factor and chromatin binding network for wood formation in *Populus trichocarpa*. *The Plant Cell, 31(3)*, 602-626.

[28] Falcioni, R., Moriwaki, T., De Oliveira, D. M., Andreotti, G. C., Souza, L. A., Santos, W. D., *et al.* (2018). Increased gibberellins and light levels promotes cell wall thickness and enhance lignin deposition in xylem fibers. *Frontiers in Plant Science, 9*, 1391.

[29] Krishnaswamy, S., Verma, S., Rahman, M. H., & Kav, N. N. V. (2011). Functional characterization of four APETALA2-family genes (*RAP2. 6*, *RAP2. 6L*, *DREB19* and *DREB26*) in *Arabidopsis*. *Plant Molecular Biology, 75(1)*, 107-127.

[30] Golz, J. F., Allen, P. J., Li, S. F., Parish, R. W., Jayawardana, N. U., Bacic, A., *et al.* (2018). Layers of regulation—Insights into the role of transcription factors controlling mucilage production in the *Arabidopsis* seed coat. *Plant Science, 272*, 179-192.

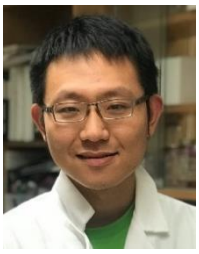
[31] Wang, Z., Yang, Z., & Li, F. (2019). Updates on molecular mechanisms in the development of branched trichome in *Arabidopsis* and nonbranched in cotton. *Plant Biotechnology Journal, 17(9)*, 1706-1722.

[32] Moore, J. P., Vicré-Gibouin, M., Farrant, J. M., & Driouich, A. (2008). Adaptations of higher plant cell walls to water loss: Drought vs desiccation. *Physiologia Plantarum, 134(2)*, 237-245.

[33] Moura, J. C. M. S., Bonine, C. A. V., Viana, J. O. F., Dornelas, M. C., & Mazzafera, P. (2010). Abiotic and biotic stresses and changes in the lignin content and composition in plants. *Journal of Integrative Plant Biology, 52(4)*, 360-376.

Haonan Tong was born May 17, 1993, Beijing, China. He received his B.S. (2015) in communication engineering from Beijing University of Post Technology, Beijing, China. His research interests include machine learning and deep learning with a keen emphasis on modelling gene regulatory mechanisms in biological systems.

Haonan is a Ph.D. candidate in the ECE Department, North Carolina State University, Raleigh, North Carolina. He is currently a research assistant on the NSF funded project "PGRP-Identification of translational hormone-response gene networks and cis-regulatory elements".



Hao Chen was born in July 14, 1990 in Yuncheng, Shandong province, China. In 2017, he received Ph.D. degree in Forest Biotechnology Group, North Carolina State University, Raleigh, North Carolina. His current research focuses on molecular mechanism plants use to integrate external signals to produce specific responses.

Hao is currently a post-doc in Alonso-Stepanova Lab, North Carolina State University, Raleigh, North Carolina. He is working on the development and implementation of next-generation-sequencing (NGS)-enabled ribosome footprinting for high-resolution whole-genome analysis of translation.

Cranos M. Williams is the director of the EnBiSys Research Laboratory in North Carolina State University, Raleigh, North Carolina. The EnBiSys Lab is a highly collaborative, multidisciplinary research laboratory, focused on the development of targeted computational and analytical solutions for modeling and controlling biological systems. Dr. Williams is the director of the Data-Driven Plant Sciences platform for the North Carolina Plant Sciences Initiative (N.C. PSI). He is well-known for his interdisciplinary research leveraging big data and machine learning to solve biological challenges.