Detecting Drought Regulators using Stochastic Inference in Bayesian Networks

Aditya Lahiri (adi441994@gmail.com)
Texas A&M University College Station https://orcid.org/0000-0001-9352-1312
Lin Zhou
Texas A&M University College Station
Ping He
Texas A&M University College Station
Aniruddha Datta
Texas A&M University College Station

Research article

Keywords: Drought, Bayesian Networks, Inference, ATAF1, MYC2

Posted Date: October 6th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-73056/v1

Abstract

Background: Drought is a natural hazard that affects crops by inducing water stress. Water stress, induced by drought, accounts for more loss in crop yield than all the other causes combined. With the increasing frequency and intensity of droughts worldwide, it is essential to develop drought-resistant crops to ensure food security. In this paper, we model multiple drought signaling pathways in Arabidopsis using Bayesian networks to identify potential regulators of drought-responsive reporter genes. Genetically intervening at these regulators can help develop drought-resistant crops.

Result: We create the Bayesian network model from the biological literature and determine its parameters from publicly available data. We conduct inference on this model using a stochastic simulation technique known as likelihood weighting to determine the best regulators of drought-responsive reporter genes. Our analysis reveals that activating MYC2 or inhibiting ATAF1 are the best single node intervention strategies to regulate the drought-responsive reporter genes. Additionally, we observe simultaneously activating MYC2 and inhibiting ATAF1 is a better strategy.

Conclusion: The Bayesian network model indicated that MYC2 and ATAF1 are possible regulators of the drought response. Validation experiments showed that ATAF1 negatively regulated the drought response. Thus intervening at ATAF1 has the potential to create drought-resistant crops.

Full Text

This preprint is available for download as a PDF.

Figures

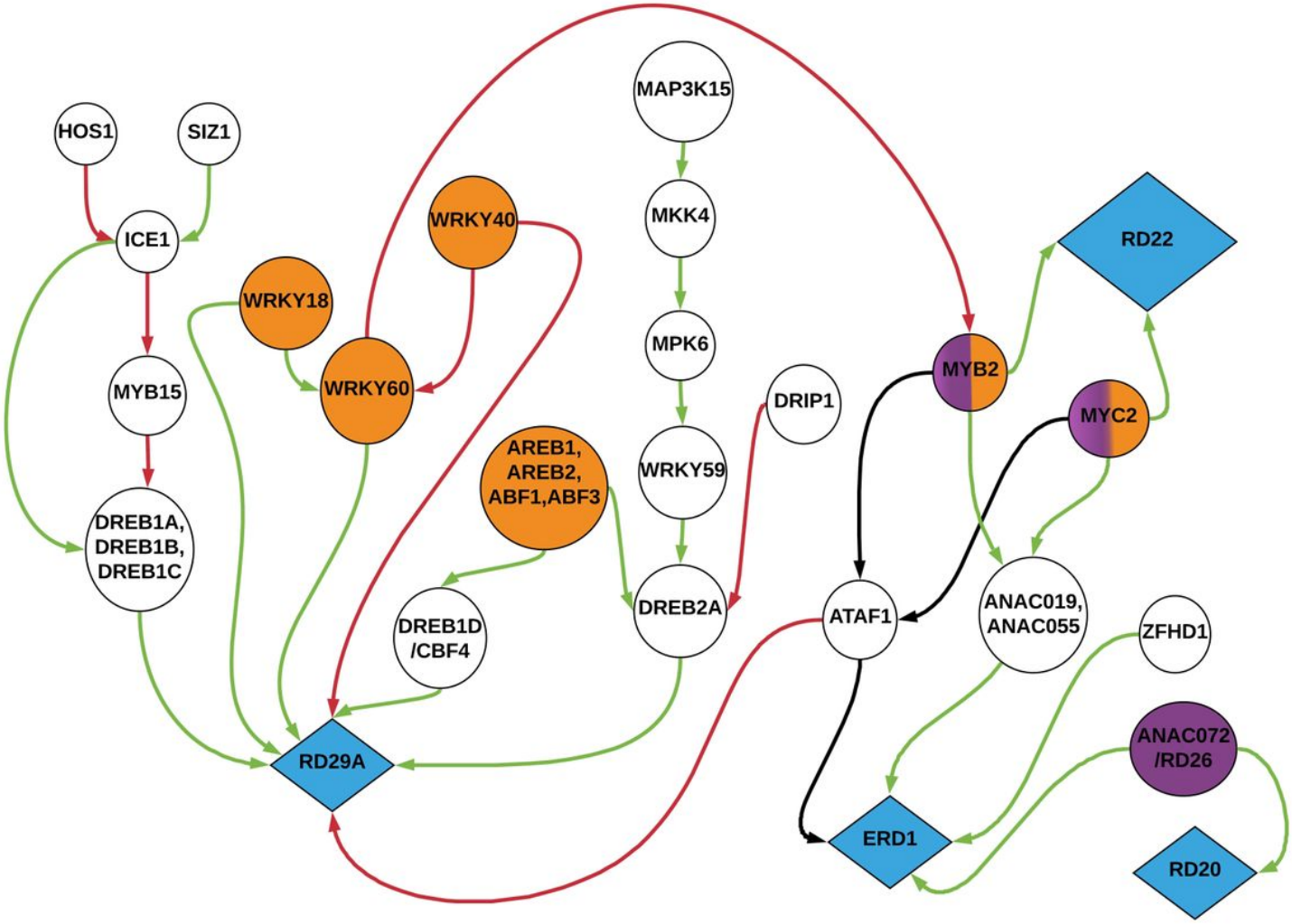

Figure 1

Drought signaling pathways in Arabidopsis. The orange circular nodes represent elements directly regulated by ABA whereas the purple nodes represent elements regulated by JA. The two nodes colored with a mix of orange and purple represent elements regulated by both JA and ABA pathways (Crosstalk). The blue diamonds represent drought responsive reporter genes. The plain circular nodes with no colors represent the transcription factors, genes and proteins involved in the regulation of drought responsive reporter genes in an ABA independent manner. The green and red arrows represent positive and negative regulation. The arrows going into and out of ATAF1 are marked black to indicate that the nature of regulation is not known at this time.

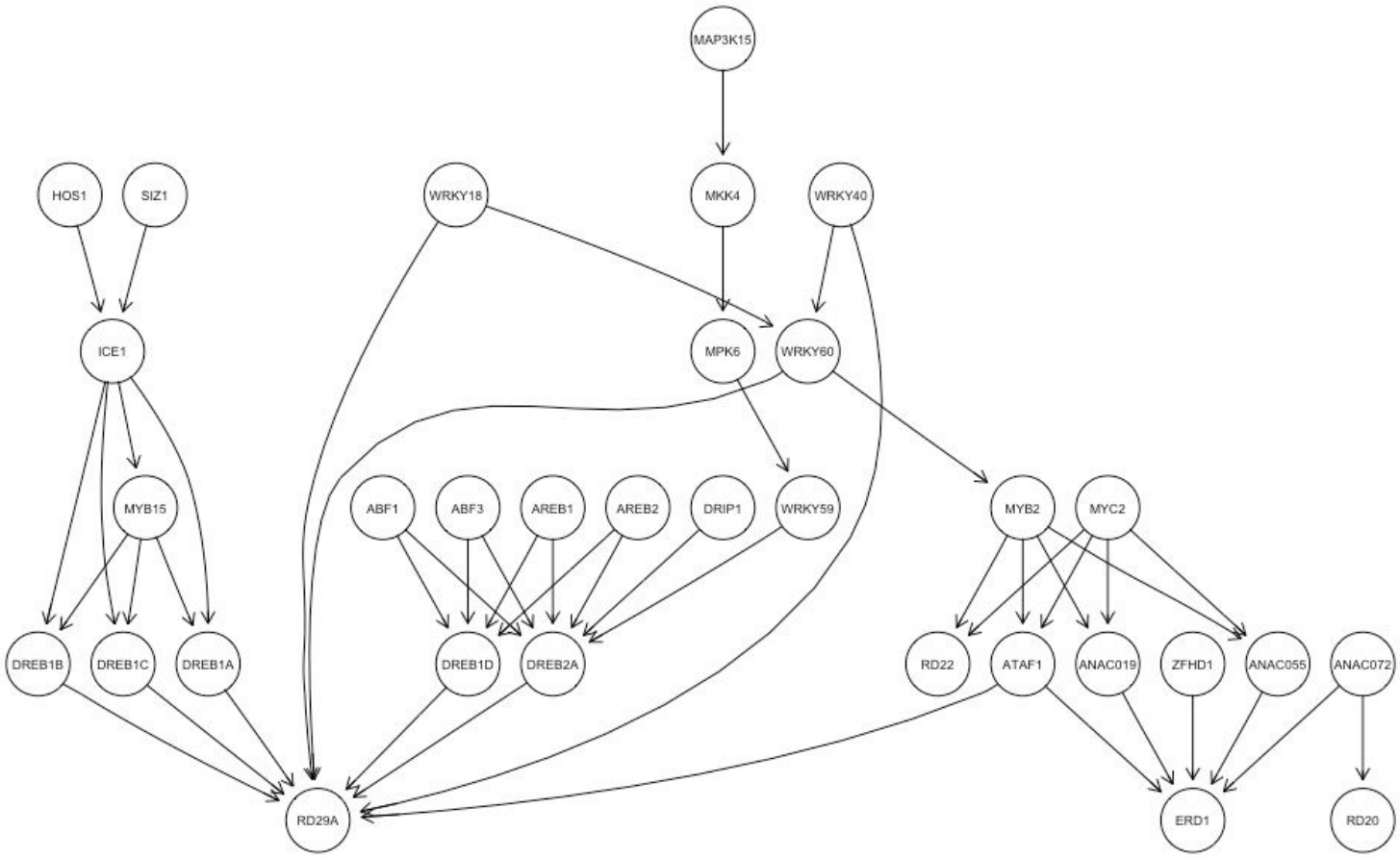

Figure 2

Bayesian Network Model of Drought Signaling Pathway. Every circular node represents a biological element in the drought signaling pathway. Every edge or black arrow represents the causal biological relationship between the nodes. Associated with every node is a θ parameter that represents the local probability distribution of the node.

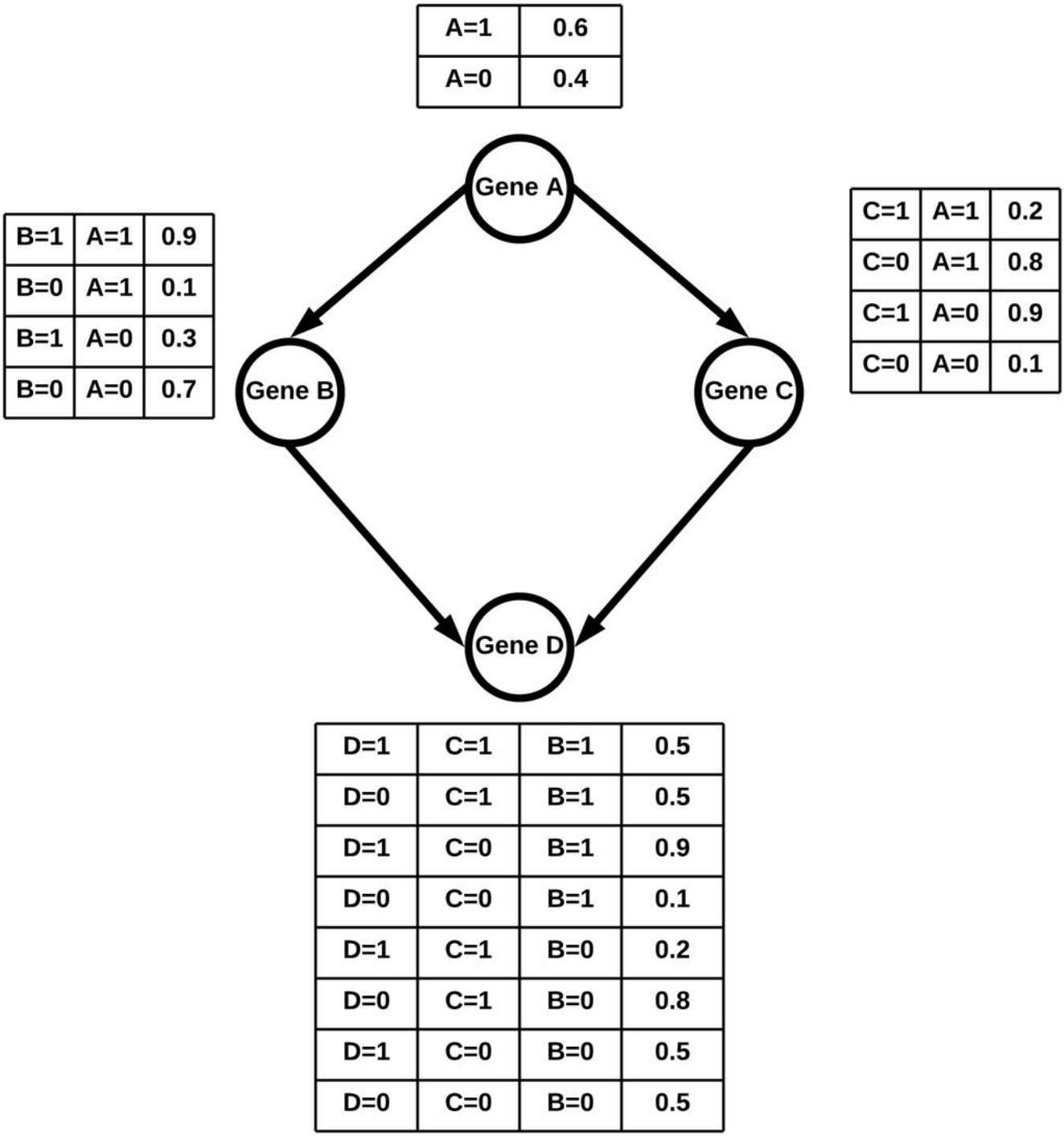

Figure 3

Example BN with LPDs. Gene A positively regulates Gene B and negatively regulates Gene C. Gene B positively regulates Gene D and Gene C negatively regulates Gene D.

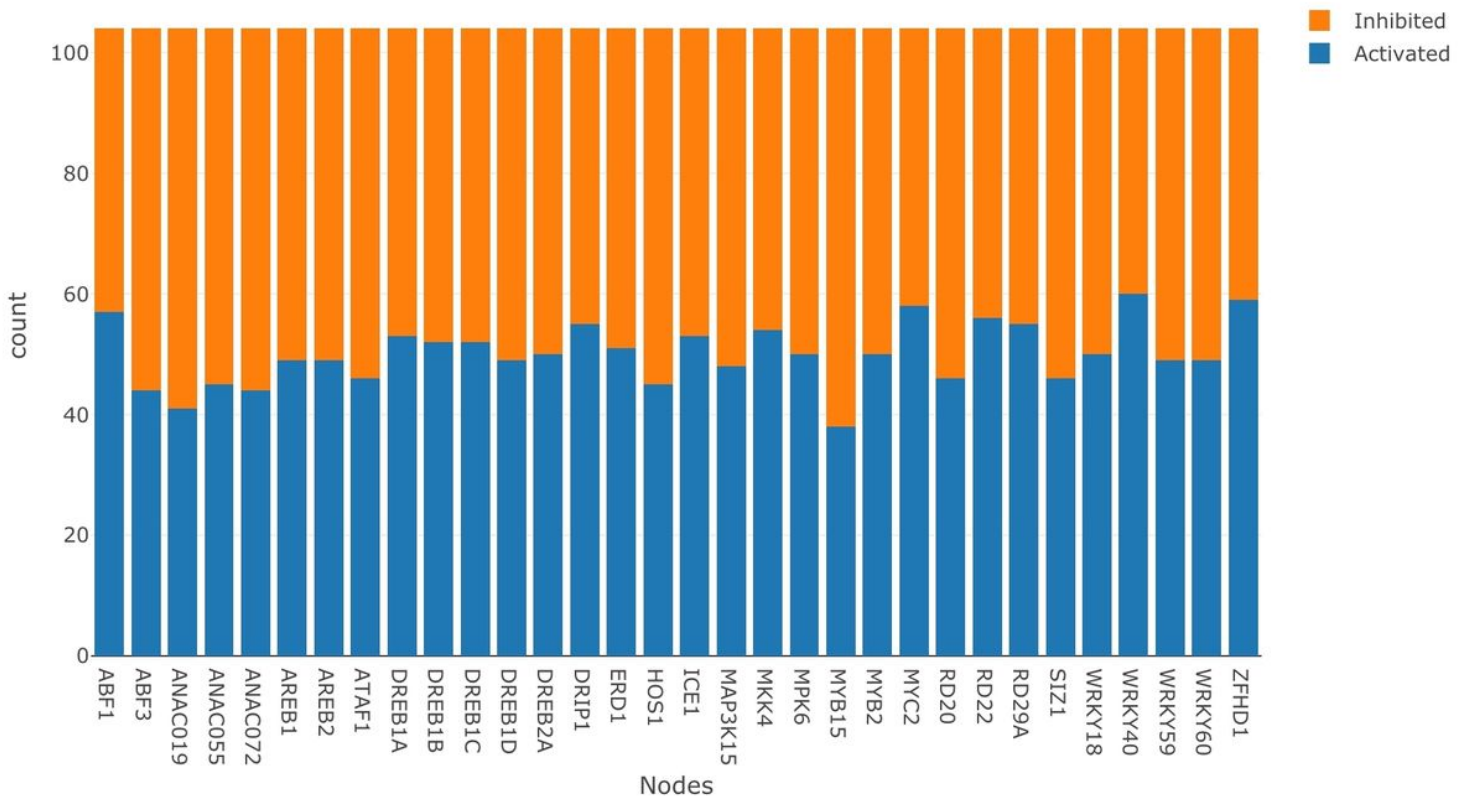

Figure 4

Activation vs Inhibition plot. This figure represents the data after it has been normalized and then binarized. There are a total of 104 data points per node. The blue part of each bar represents activation counts whereas the orange part represents the inhibition counts.

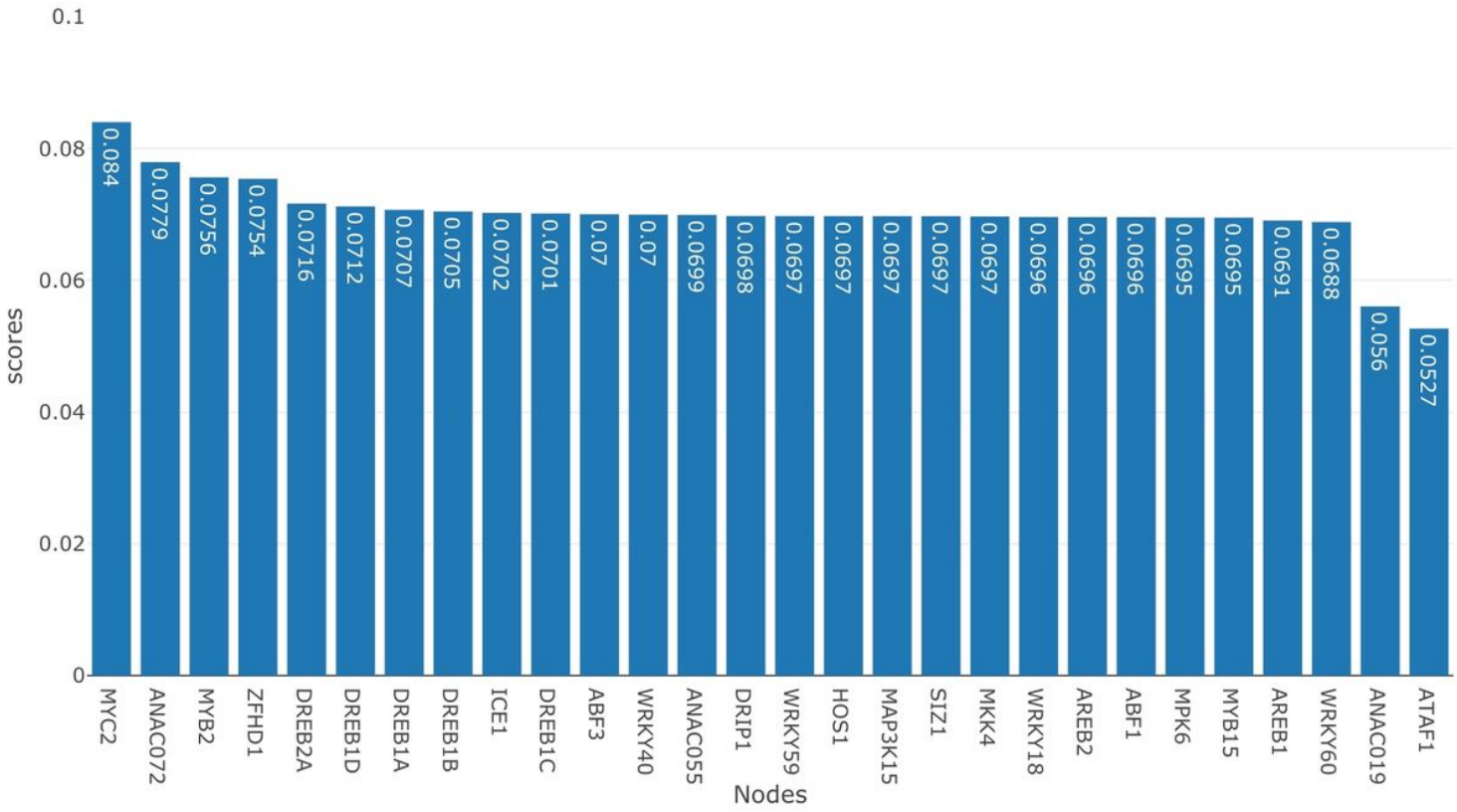

Figure 5

Activation Scores for non-reporter gene nodes. Associated with each node is a blue bar which represents the score for activating that node.

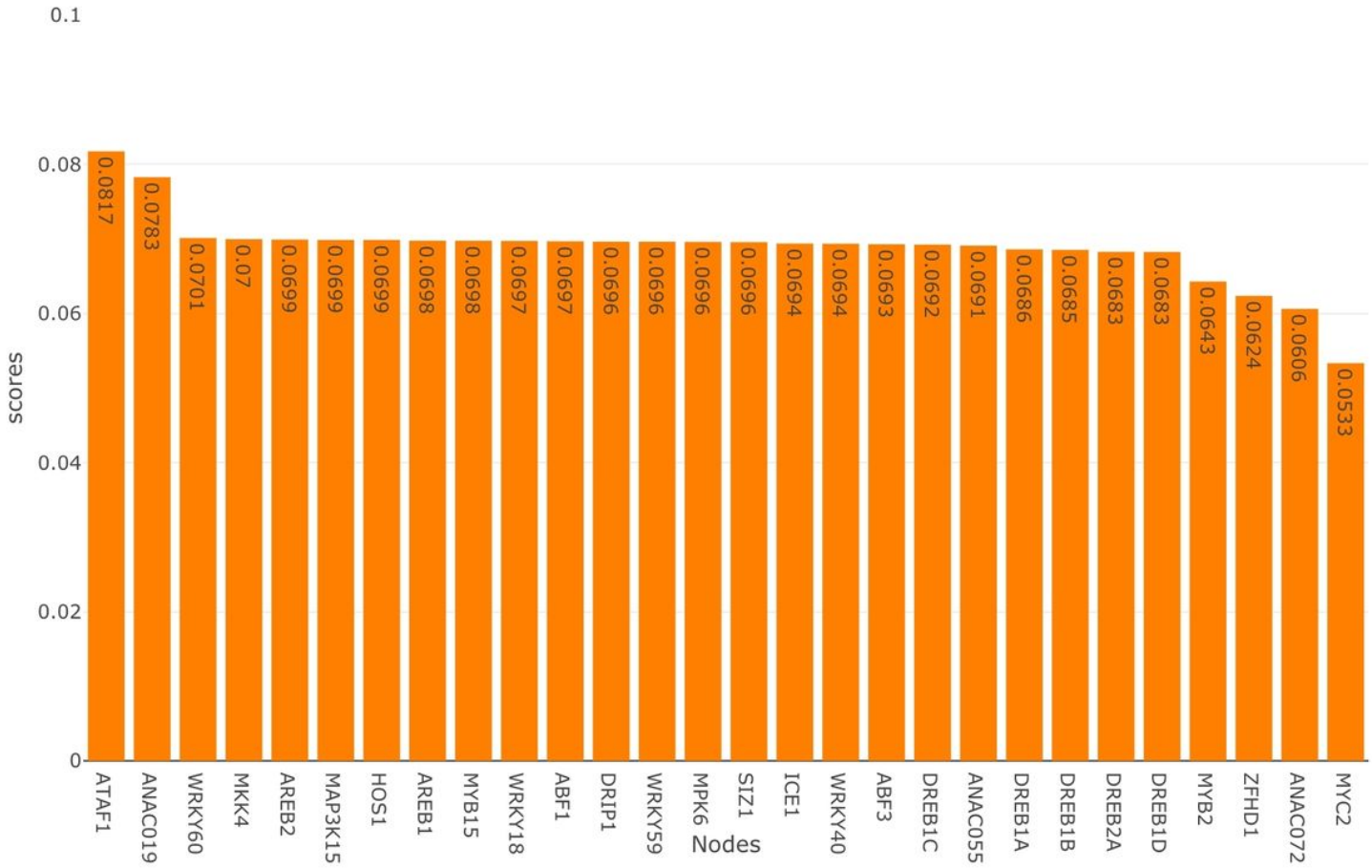

Figure 6

Inhibition Scores for non-reporter gene nodes. Associated with each node is an orange bar which represents the score for activating that node.

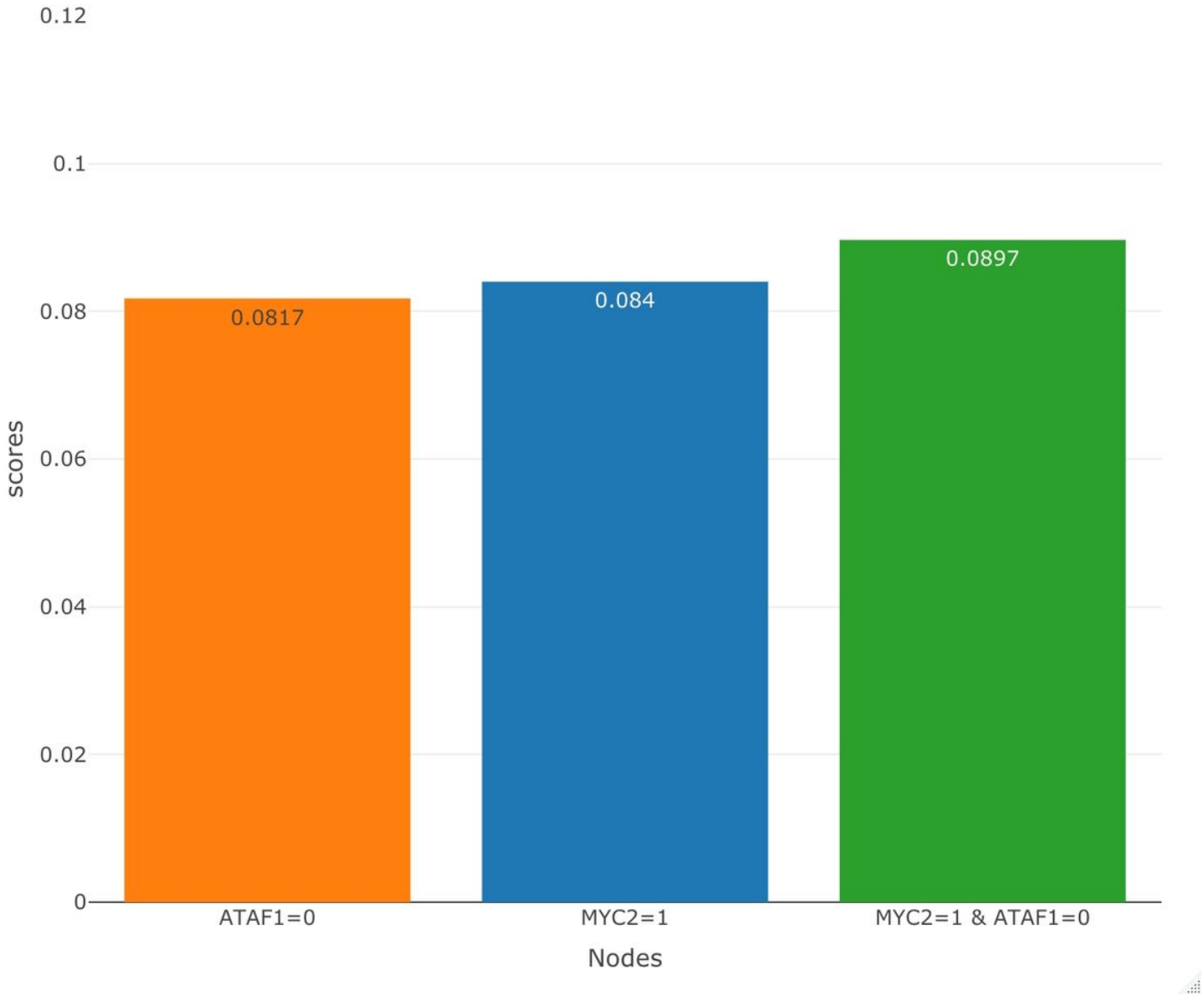

Figure 7

Comparing the scores of multi-node and single node intervention under optimal response case. Simultaneous (multi-node) intervention on MYC2 and ATAF1 has a slightly higher score than single node intervention.

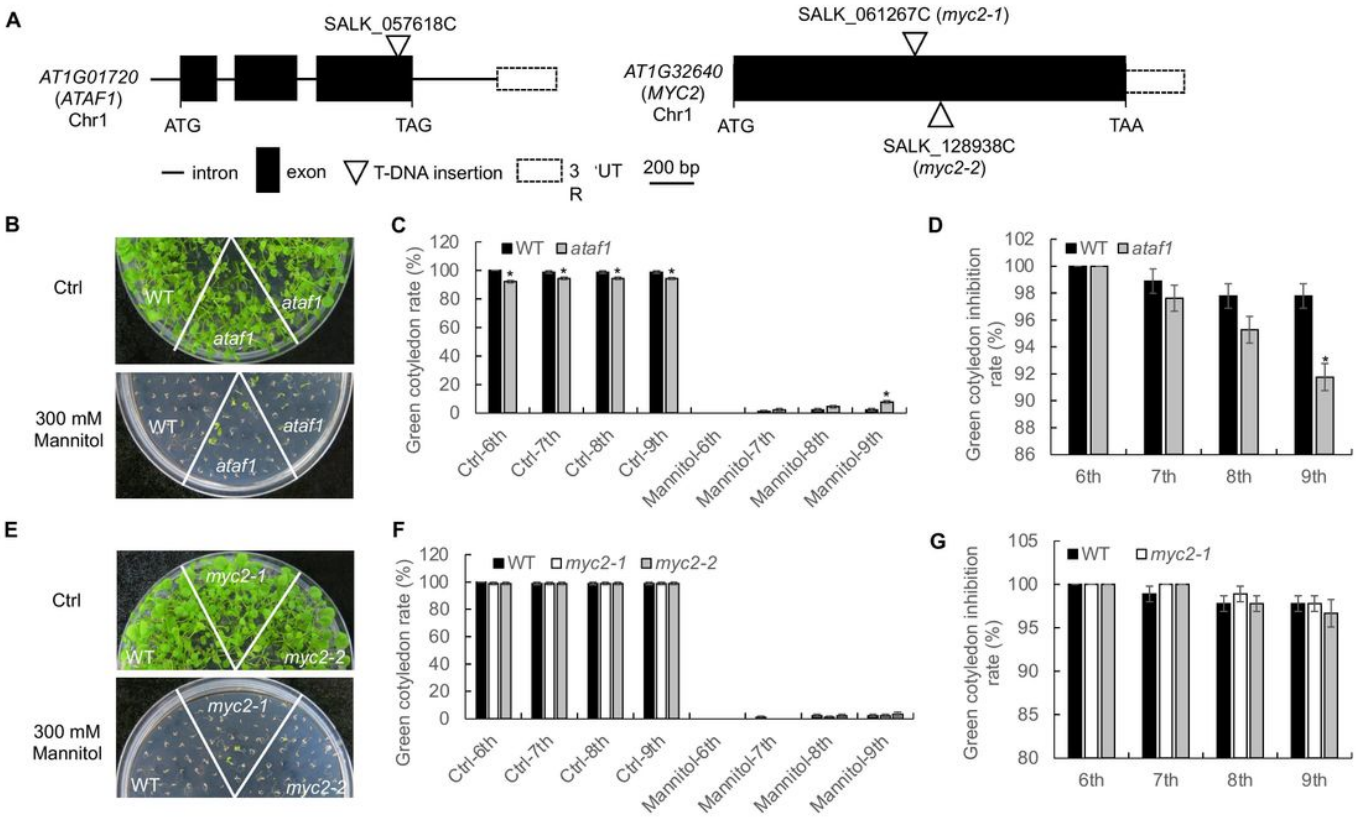

Figure 8

Results from validation experiments. A.The scheme of the ATAF1 and MYC2 genomic DNA and T-DNA insertion. The panel is a schematic illustration of the ATAF1 and MYC2 genomic DNA with exons (solid box), intron (lines) and 3' untranslated region (open box). The position of T-DNA insertion of ataf1 (SALK 057618C), myc2 (SALK 061267C, SALK 128938C)was labeled. B.The ataf1 mutant is more resistant to mannitol treatment. Wild-type (WT) Col-0 and ataf1 mutant seeds were germinated on 1/2 MS medium with or without 300 mM mannitol. 30 seeds per genotype were used for each replicate. The photos were taken four-week post-germination. C.Quantification of cotyledon greening on plates corresponding to B. Seedlings with green cotyledon expansion were counted at 6-9 days post-germination. Data are shown as means ± SD (standard deviation) from three independent replicates (n=3, *, p¡0.05, Student's t-test). D.Quantification of cotyledon greening inhibition rate on plates corresponding to B. Seedlings with green cotyledon expansion were counted at 6-9 days post-germination. Data are shown as means ± SD from three independent replicates (n=3, *, p¡0.05, Student's t-test). E.Growth of WT and myc2 mutants on MS plates. WT and myc2 mutant seeds were germinated on 1/2 MS medium with or without 300 mM mannitol. 30 seeds per genotype were used for each replicate. The photos were taken four-week post-germination. F.Quantification of cotyledon greening on plates corresponding to E. Seedlings with green cotyledon expansion were counted at 6-9 days post-germination. Data are shown as means ± SD from three independent replicates (n=3, no statistical significance with Student's t-test). G. Quantification of cotyledon greening on plates corresponding to E. Seedlings with green cotyledon expansion were counted at 6-9 days post-germination. Data are shown as means ± SD from three independent replicates (n=3, no statistical significance with Student's t-test).