

## Learning MAPK Pathway from Data

MAPK pathway or (Mitogen-Activated Protein Kinase) pathway is a well-known signaling pathway in plants that is triggered under biotic stress. In plants it plays a critical role in deploying defense response against microbial pathogens. Various biological studies have implicated the MAPK pathway as a point of convergence under biotic stresses [1]. Hence it is useful to study this network for understanding the role this plays at the genetic level.

In this project, we will use the dataset **GSE17464**, **GSE19109**, and **GSE18978** which are publicly available from NCBI GEO to learn the MAPK signaling pathway. The MAPK signaling pathway as explained in biological literature is shown in figure 1 [2].

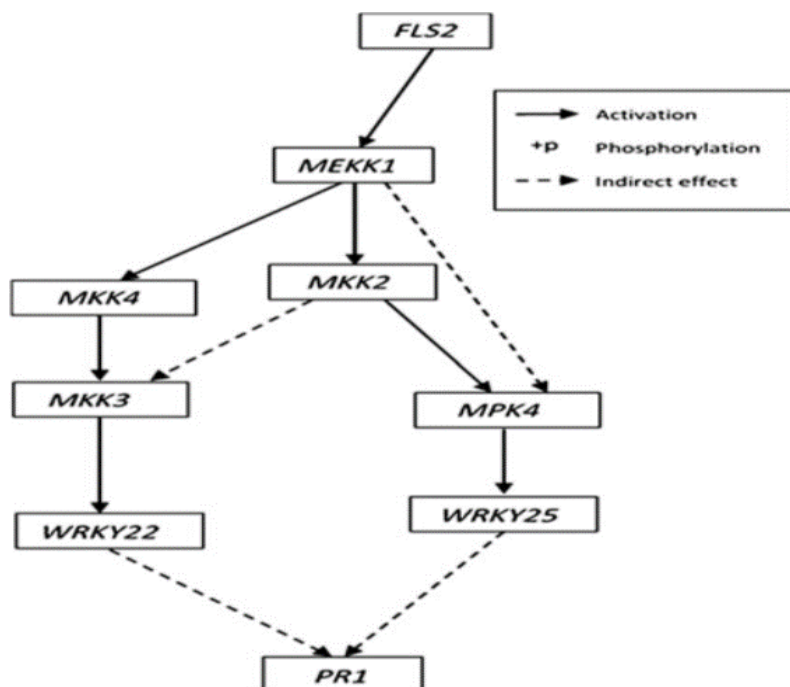


Figure 1: MAPK Signaling pathway converging to defense response gene PR1

Since the MAPK network is a causal network, we will use these datasets **to derive a Bayesian Network model** that should capture the genetic interactions that exist among the elements of the MAPK network. So our overall objective will be to estimate how close the derived Bayesian network is to the actual biological pathway that is described in figure 1. Bayesian networks will be ideal for this project as every node of the Bayesian network will serve as a gene in the MAPK pathway while every directed edge will serve to describe the causal interactions that exist among each of the genes. To learn the structure of the Bayesian network we will first assume a complete graph and then use the Chi Squared statistics as a deviance measure to determine if an edge should exist between two nodes. To direct the network we will evaluate and implement the immoralities and then apply the rules for deriving class PDAG to derive the overall Bayesian network. We will assess the accuracy of the final Bayesian network by evaluating the true positives, true negatives, false

positives and false negative edges. One of the drawbacks of using this method is that we may encounter undirected edges in our derived Bayesian network. Since our Bayesian network is essentially a PDAG and it may contain undirected edges which may lead to multiple possible graph structures.

### **REFERENCES**

- [1] Brader, G., Djamei, A., Teige, M., et al. 2007. The MAP kinase kinase MKK2 affects disease resistance in Arabidopsis. *Mol. Plant Microbe Interact.* 20, 589–596.
  
- [2] Venkat, P., Narayanan, K. and Datta, A. (2017). A Bayesian Network-Based Approach to Selection of Intervention Points in the Mitogen-Activated Protein Kinase Plant Defense Response Pathway. *Journal of Computational Biology*, 24(4), pp.327-339.