IDENTIFICATION OF KEY REGULATORS OF GUARD CELL FUNCTIONS USING SYSTEMS BIOLOGY APPROACHES

A Dissertation in
Integrative Biosciences

by

Song Li

© 2010 Song Li

Submitted in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

December 2010
The dissertation of Song Li was reviewed and approved* by the following:

Sarah M. Assmann  
Waller Professor of Biology  
Dissertation Advisor  
Co-chair of Committee

Réka Albert  
Professor of Physics and Biology  
Co-chair of Committee

Andrew G. Stephenson  
Professor of Biology and Assistant Department Head for Research

Claude dePamphilis  
Professor of Biology

Debashis Ghosh  
Professor of Statistics

Peter Hudson  
Willaman Professor of Biology  
Director, Huck Institutes of the Life Sciences

*Signatures are on file in the Graduate School
ABSTRACT

Over the past decade, fully sequenced genomes and the advent of the high throughput technologies such as transcriptome profiling with microarrays, have provided biologists catalogs of the molecular components of many biological systems. While traditional genetic and molecular biology approaches focus on identifying functions of one individual component at a time, systems biology approaches have become essential for biologists to understand the complex interactions of the cellular components, and to distill key regulators out of the rich data that are generated by high throughput experiments. I have undertaken two projects that address two general questions in systems biology: 1) how to reconstruct the network of the interacting molecular components, and to simulate the cellular process that is carried out by the molecular network. 2) Given a compendium of gene expression data from multiple tissues in a multicellular organism, how to identify novel genes that function in a specific cell type. In both projects, I have used plant guard cells as a model system, while each project has been developed using specific computational methods.

Dozens of cellular components have been found to be involved in the guard cell ABA response, however, most interactions between these components are indirect, and the detailed parameters of known direct reactions are largely unknown. A Boolean network approach is adopted to address the problem of lack of details regarding the guard cell signal transduction network. Random asynchronous update and random initial states of the components are introduced to model both the uncertainty that is inherent in the intracellular signaling process as well as our incomplete knowledge of the signaling process. Simulations are carried out to determine whether knockout of each component has an impact on the ABA induced stomatal closure. Computational predictions are validated by wet-bench experiments using chemical inhibitors that “clamp” intracellular signaling components.
In a multicellular organism, each cell type performs specialized functions using genes that are expressed in that cell type. To understand the general principle of how genes are expressed in multicellular organisms, tissue-based gene expression patterns are analyzed for thousands of genes from four multicellular organisms: human, mouse, rice and Arabidopsis. Expression kurtosis is found to be the common organizing principle of tissue preferential expression in all four organisms, thus is used to define in which tissues a high kurtosis gene is preferentially expressed, for every high kurtosis genes in both human and Arabidopsis. To identify novel genes that are essential to guard cells’ functions, transcriptome data are obtained using Arabidopsis transcriptome microarray. A linear model approach is used to merge the expression profiles generated by our group with published results from other published transcriptome data before the identification of high kurtosis genes that are preferentially expressed in guard cells. Many high kurtosis genes are found to be preferentially expressed in multiple tissues. Using high kurtosis genes as signatures for tissue functions gene-centric tissue networks are constructed to explore the similarities between different tissues. Tissues from similar organ origination are found to group into modules in the gene-centric tissue networks, supporting the predictive power of the tissue network approach. Four genes that are newly predicted to function in guard cells and/or roots are validated using knockout mutants, and the phenotypes of mutants agree well with the in silico predictions.

The results of gene tissue associations and the tissue networks are of general interest to biologists. A web interface that allows interactive visualization and exploration of the gene centric tissue networks is available at http://personal.psu.edu/szl116/TissueNetworks.
TABLE OF CONTENTS

LIST OF FIGURES ........................................................................................................... viii

LIST OF TABLES ............................................................................................................ x

ACKNOWLEDGEMENTS ................................................................................................. xi

Chapter 1 Introduction ................................................................................................. 1

1.1 Introduction to computational systems biology ..................................................... 1
1.2 Introduction to guard cells: a model system for plant systems biology ............... 1
1.3 Bottom-up approach of systems biology: dynamic modeling of the guard cell signaling network .................................................................................................................. 2
1.4 Top-down approach of systems biology: high kurtosis genes in multicellular organisms and identification of novel genes that function in guard cells. .............. 4
1.5 Dissertation Outline ............................................................................................... 6

Chapter 2 Genetic determinants of stomatal function ................................................. 7

2.1 Introduction ........................................................................................................... 7
2.2 Arabidopsis as a model system .............................................................................. 10
2.3 How do stomates sense drought stress .................................................................. 11
2.3.1 ABA as a drought hormone ........................................................................... 12
2.3.2 Other hormones in stomatal drought responses ............................................. 14
2.4 Signaling events inside guard cells in response to drought ................................. 16
2.4.1 Ion channels and other membrane transport related proteins ...................... 18
2.4.2 Calcium ......................................................................................................... 20
2.4.3 ROS and NO ................................................................................................. 21
2.4.4 Other small intracellular molecules ................................................................ 21
2.5 Cell signaling mutants with altered stomatal responses ....................................... 22
2.5.1 Protein Kinases ............................................................................................ 22
2.5.2 Phosphatases ................................................................................................ 27
2.5.3 G proteins .................................................................................................... 30
2.5.4 Farnesyl Transferase .................................................................................... 34
2.5.5 Genes related to humidity sensing ............................................................... 35
2.6 Transcriptional regulation in stomatal drought response .................................... 36
2.6.1 Transcription factors in stomatal drought response .................................... 36
2.6.2 Stomatal development and drought responses ............................................. 38
2.7 Summary ............................................................................................................... 39

Chapter 3 Predicting essential components of signal transduction networks: a dynamic model of guard cell abscisic acid signaling ..................................................... 42

3.1 Abstract ............................................................................................................... 42
3.2 Introduction .......................................................................................................... 43
3.3 Results ................................................................................................................... 46
3.3.1 Extraction and Organization of Data from the Literature ............................... 46
3.3.2 Assembly of the ABA Signal Transduction Network ................................... 48
3.3.3 Modeling ABA Signal Transduction......................................................... 53
3.3.4 Identification of Essential Components.................................................... 61
3.3.5 Experimental Assessment of Model Predictions........................................... 69
3.4 Discussion ......................................................................................................... 72
3.4.1 Network Synthesis and Path Analysis......................................................... 72
3.4.2 Dynamic Modeling ........................................................................................ 74
3.4.3 Importance of $\text{Ca}^{2+}$ Oscillations to ABA-Induced Stomatal Closure ...... 77
3.4.4 Limitations of the Current Analysis ............................................................. 79
3.4.5 Implications ................................................................................................... 81
3.6 Materials and Methods................................................................................... 84

Chapter 4 High kurtosis is an essential principle of gene expression in multicellular organisms ................................................... 86
4.1 Abstract.............................................................................................................. 86
4.2 Introduction ...................................................................................................... 87
4.3 Results................................................................................................................ 92
4.3.1 Kurtosis of gene expression patterns in four multicellular organisms .......... 92
4.3.2 Gene tissue associations and gene tissue network analysis ......................... 93
4.3.3 Functional analysis of leptokurtically distributed genes associated with guard cells .................................................................................................................. 105
4.3.4 Orthologs of high kurtosis genes are also high kurtosis genes in the counterpart organism ................................................................................................. 112
4.4 Discussion ......................................................................................................... 114
4.4.1 A new organizing principle of gene expression in multicellular organisms... 114
4.4.2 Gene-tissue co-association is complementary to gene co-expression analysis .................................................................................................................... 115
4.4.3 High kurtosis genes and the evolution of multicellularity ............................ 116
4.4.4 Tissue networks represent higher dimension genome expansion ............... 118
4.5 Materials and Methods................................................................................... 120

Chapter 5 Summary and Future Directions ........................................................................ 122
5.1 Summary .......................................................................................................... 122
5.2 Future research interests .................................................................................. 124
5.2.1 Extension to the bottom-up modeling approach. ........................................ 124
5.2.2 Extension to the top-down approach. ......................................................... 125

Appendix A. Supplemental Information for Chapter 3 ................................................ 127
Appendix A. Figure S1: Probability of closure in randomized networks where pairs of positive (A) or negative (B) edges are rewired ............................................. 128
Appendix A. Table S1. Synthesis of Experimental Information about Regulatory Interactions between ABA Signal Transduction Pathway Components ............ 129
Appendix A. Text S1: Detailed Justification for Each Boolean Transfer Function .... 134
Appendix A. Text S2: Verification of the Inference Process and of the Resulting Network .................................................................................................................... 141
Appendix A. Text S3: The Effect of Random Rewiring on the Network Dynamics ..... 142
LIST OF FIGURES

Figure 2.1. Stomatal drought response is mediated by plant hormones. ......................... 12
Figure 2.2. Cellular mediators of stomatal drought response........................................ 17
Figure 3.1. Illustration of the Inference Rules Used in Network Reconstruction. .......... 48
Figure 3.2. Current Knowledge of Guard Cell ABA Signaling...................................... 50
Figure 3.3. Stomatal Aperture Distributions without ABA Treatment (gray bars) and with 50 µM ABA (white bars)........................................................... 57
Figure 3.4. Schematic Illustration of Our Modeling Methodology and of the Probability of Closure. ........................................................................................................ 60
Figure 3.5. The Probability of ABA-Induced Closure as a Function of Timesteps in the Dynamic Model. ...................................................................................... 65
Figure 3.6. Classification of Close-to-Normal Responses........................................... 67
Figure 3.7. Summary of the Dynamic Effects of Calcium Disruptions. ...................... 69
Figure 3.8. Effect of Cytosolic pH Clamp on ABA-Induced Stomatal Closure. .......... 70
Figure 4.1. Distribution of gene expression kurtosis in four multicellular organisms. .... 91
Figure 4.2. Human gene-centric tissue network......................................................... 96
Figure 4.3. Arabidopsis gene-centric tissue network ................................................... 97
Figure 4.4. Graph theoretical properties of tissue networks with Q = 1e-09 and Z = 10 and different S values................................................................. 98
Figure 4.5. Median numbers of leptokurtic genes in each module of the human and Arabidopsis tissue network................................................................. 100
Figure 4.6. Arabidopsis shoot apex subnetwork, Arabidopsis seed subnetwork and human liver subnetwork............................................................... 104
Figure 4.7. Stomatal phenotypes of mutants of four guard cell leptokurtic genes.......... 109
Figure 4.8. Disease genes are over-represented in human leptokurtic genes. ............... 111

Figures in Appendix A:
**Figure S1.** Probability of closure in randomized networks where pairs of positive (A) or negative (B) edges are rewired. ................................................................. 128

**Figures in Appendix B:**

**Figure S1.** Reduction of kurtosis plots................................................................. 175

**Figure S2.** Clustering profiles of R12 to R45 of human and Arabidopsis HKGs.......... 176

**Figure S3.** Tissue associations of leptokurtic genes derived by our method perform better than gene-tissue associations from other published methods.......................... 177

**Figure S4.** Graph theoretical properties of tissue networks with Q = 0.005 and Z = 3........ 177

**Figure S5.** GO annotation analysis for genes appearing in seed subnetwork............ 179

**Figure S6.** GO annotation analysis for genes appearing in shoot apex subnetwork .......... 181

**Figure S7.** GO analysis for human liver and fetal liver subnetwork.......................... 183

**Figure S8.** Supporting experimental results......................................................... 185

**Figure S9.** Root phenotypes of T-DNA insertional mutants of At1g11100.................... 186
LIST OF TABLES

Table 2.1. Mutants with both stomatal phenotypes and increased drought tolerance .......... 9

Table 3.1. Boolean Rules Governing the States of the Known (Named) Nodes in the Signal Transduction Network ................................................................. 58

Table 3.2. Single to Triple Node Disruptions in the Dynamic Model ............................... 63

Table 4.1. First neighbor tissues of guard cells.............................................................. 106

Table 4.2. GO analysis for human and Arabidopsis HKGs .......................................... 113

Tables in Appendix A:

Table S1. Synthesis of Experimental Information about Regulatory Interactions between ABA Signal Transduction Pathway Components. ........................................ 129

Table S2. Perturbation analysis of Table S1 .............................................................. 141

Tables in Appendix B:

Table S1. Tissue annotations for Figure 4.2 ............................................................. 187

Table S2. Tissue annotations for Figure 4.3 ............................................................. 189

Table S3. HKGs with known functional roles in guard cells ..................................... 193

Table S4. Gene Ontology (GO) analysis of leptokurtic genes in guard cells and neighboring tissues ................................................................. 194

Table S5. Primers used for real time PCR analysis of GC-abundant genes and for genotyping of knock-out lines ................................................................. 197
ACKNOWLEDGEMENTS

I would like to first thank my advisors, Dr. Sarah Assmann and Dr. Réka Albert, their patience, and financial support during my time here. I also feel very grateful to the other committee members, Dr. Debashis Ghosh, Dr. Claude dePamphilis, and Dr. Andrew G. Stephenson. I thank them for serving on my dissertation committees and providing valuable suggestions on my dissertation work. I would like to specially thank people that contribute to the microarray data: Dr. Sona Pandey, Timothy Gookin, Dr. Zhixin Zhao and Liza Wilson. I would also like to thank many people from Réka’s lab: Dr. Jaewook Joo, Dr. Juilee Thakar, Dr. Ranran Zhang, Dr. Ruisheng Wang, and Assieh Saadatpour, for their valuable discussion on my project.

Most of all, I would like to thank my wife Yao Zhang, for her constant support, love and understanding through all these years; my parents and in-laws who provided constant encouragement and support, and for taking care of my son when I am busy with my research work.
Chapter 1 Introduction

1.1 Introduction to computational systems biology

Biological systems are characterized by the interconnected networks of varieties of components: DNA, RNA, proteins, other organic compounds and inorganic ions. Knowing the properties of the components of such complex systems usually cannot directly translate into an understanding of the behavior of the whole system, e.g. the developmental program of an organism or the responses of an organism to environmental stimuli. Systems biology aims to reveal the properties of complex biological systems through integrating genome scale data of the cellular components as well as the connections and regulations between these components. Mathematical modeling of biological networks and data mining of genome scale datasets have become two interwoven disciplines of systems biology towards the non-reductionist view of living organisms. The reminder of this chapter provides an overview of the applications of both approaches to advance our understanding of a particular plant cell type—guard cells.

1.2 Introduction to guard cells: a model system for plant systems biology.

The ultimate goal of systems biology is to understand biology; hence, any in silico prediction from a systems biology approach has to be empirically tested in a real biological system in order to assess the predictive power of that approach. Here we choose to use guard cells as our model in plants. Arabidopsis guard cells are an ideal model system for plant systems biology, because: 1) the genome of Arabidopsis has been fully sequenced and is the best annotated plant genome to date [1, 2]. 2) A large number of genetic regulators of key biological processes have been identified through traditional genetics, while gene knockouts of most genes
with or without functional annotations are available [3]. 3) More than 40 genes functioning in
guard cells, as well as many small molecules have been identified over the past decades [4, 5].
The existing knowledge and experimental resources are crucial for us to apply systems biology
approaches in guard cells.

To create a predictive and systems level model of guard cells, one needs to first
summarize the known regulators of guard cells’ functions. Pairs of guard cells surround stomatal
pores, which can open and close in response to water availability, light intensity, and the carbon-
dioxide concentration in the air. Guard cells play pivotal roles in plant drought tolerance through
response to the plant hormone abscisic acid (ABA). ABA accumulates under drought stress and
triggers the guard cell signaling network, which is made of a plethora of signaling molecules,
such as protein kinases, phosphatases, and G proteins. The intracellular signaling events lead to
the activation or the inhibition of the membrane localized ion channels and ion pumps, which
mediate the changes of cellular solute concentration, and eventually, cause the changes of the
volume of guard cells and the stomatal apertures. Genes and intracellular processes that are
related to guard cell ABA responses are summarized in chapter 2.

1.3 Bottom-up approach of systems biology: dynamic modeling of the guard cell signaling
network

Traditional approaches have discovered dozens of intracellular components that are
involved in guard cell ABA signaling. The complexity of the interacting patterns of these
components makes the system responses unpredictable by human intuition alone. Computational
modeling of signaling networks is the inevitable route for us to gain predictive understanding of
guard cell ABA responses, to generate rationale hypotheses, and to design experiments that aim
to further elucidate the functions of guard cells. Computational and mathematical modeling of a biological system with known components in the system and known interactions between the known components in the system is a bottom-up approach of systems biology.

Mathematical models have been widely used to explain biological phenomena, with different types of the mathematical models that emphasize different details of the system under study [6]. One example is the use of networks and graph theory to model protein-protein interactions [7, 8]. Proteins with drastically different molecular weights and sizes are all represented as “nodes” in graph theoretical terms, and interactions between proteins with different underlying mechanisms are represented by “edges” between nodes. While details regarding the sizes, structures, and stoichiometries of interacting proteins have been ignored in these analyses, network models have lead to testable hypotheses and systems level insight into the organization of proteins inside living cells [9]. Another example of mathematical models is to model a signaling transduction process as a system of chemical reactions, using fully stochastic simulation of all molecular interacting events [10, 11]. Such models, usually limited by both the lack of detailed knowledge of reactions in a cellular system and the lack of kinetic parameters, can only be successful applied in small scale systems that are composed of a handful of signaling proteins.

The essence of making a good model that addresses particular questions of a biological system is multifold. The model has to incorporate the interactions of the components in a system, because the interactions are essential to the system-level function, which is more than the sum of the functions of the individual components. One virtue of mathematical models is that the computational power allows us to explore the responses of a biological system under conditions that are yet untested by experiments. However, some predictions made by a model have to be amenable for experimental tests given the available technical resources, such that at least some predictions can be validated. A useful model usually takes a certain level of abstraction of the
biological system, such that the existing knowledge of model parameters or structures are sufficient for the model to make testable predictions, while neglecting other details. In this study, we start by assembling the signaling networks from published articles of ABA signaling in guard cells, and then construct a Boolean dynamic model for the process of ABA induced stomatal closure. The model prediction and experimental validation are discussed in Chapter 3.

1.4 Top-down approach of systems biology: high kurtosis genes in multicellular organisms and identification of novel genes that function in guard cells.

From the model of ABA induced stomatal closure, we find that many interactions between known components are indirect interactions, e.g. one protein kinase could be the positive regulator of the production of a small signaling molecule, but the protein kinase does not function directly in catalyzing the formation of the small signaling molecule. One possibility is that these indirect interactions are mediated by novel genes that have not been identified by traditional approaches. We also find that there are several seemingly independent paths in the ABA signaling network. These independent paths may cross talk with each other, through yet unidentified gene products. Therefore, a method that allows discovery of novel components involved in a cell signaling pathway is of great interest to the understanding of guard cells, as well as to many other less characterized biological systems.

The fully sequenced genomes of *Arabidopsis* and many other model organisms have enabled the explosive development of high-throughput approaches, which provide quantitative measures for large numbers of intracellular components with unprecedented speed and coverage. Datasets from transcriptome, proteome and metabolome have fueled the top-down approaches of systems biology, i.e. identification of novel components that function in a system by mining genome scale data. For example, microarrays now have been routinely used to investigate
expression changes of tens of thousands of transcripts in samples under different external stimuli or different developmental stages of tissue and organs. Differentially expressed genes under these conditions or developmental stages are further investigated as candidate genes that are likely to participate in condition-specific cellular networks.

Computational and statistical analyses are necessary steps for distilling signals out of noise from large scale datasets. However, a good “top-down” approach usually distinguishes itself from other approaches in the sense that the underlying “organizing principles” of biological systems can be revealed [12, 13]. Such “organizing principles” cannot be identified and validated without genome wide and, sometimes, cross species comparison of genome scale data. These organizing principles can significantly increase our understanding of biological systems, and also provide practical guidance for applications such as manipulating traits using transgenic approaches in plants [14] and identifying genes that are associated with disease processes in humans [15].

To this end, we carry out transcriptome analysis of guard cells using the *Arabidopsis* full genome microarray (ATH1), which can simultaneously measure the transcript abundance of more than 20,000 genes in the *Arabidopsis* genome. We are seeking a general “organizing principle” of how genes are differentially expressed in different tissues of multicellular organisms. We find kurtosis is the “organizing principle” for tissue preferential expression in four multicellular organisms: human, mouse, rice and *Arabidopsis*. We also find that the majority of high kurtosis genes are “shared” between tissues types, and can be used as the descriptors for the functions of the tissues that they are expressed in. Finally we create gene-centric tissue networks which allow the exploration and identification of genes involved in different tissue functions in human and *Arabidopsis*. These results are further discussed in Chapter 4.
1.5 Dissertation Outline

This dissertation consists of a review of the genetic regulators of guard cells functions, and two projects covering both the bottom-up and top-down approaches of systems biology of guard cells.

In chapter 2, a survey of the published genes involved in guard cell signaling is presented, which was originally published as a book chapter. In chapter 3, a dynamic model of guard cell signaling is presented, and this model was published as a journal article. In chapter 4, the high kurtosis genes and tissue networks are discussed. This project represents two manuscripts in preparation to submit to PNAS and an opinion journal. Chapter 5 draws conclusions from the results of these projects and describes ideas for future directions.
Chapter 2 Genetic determinants of stomatal function

This chapter is previously published as a book chapter [16] and was written specifically to fulfill thesis requirements.

2.1. Introduction

Water is the major constituent of land plants, comprising 80-90% of the fresh weight of most herbaceous plants [17]. As a small polar molecule with a high dielectric constant, water is a good general purpose solvent for many organic and inorganic ions. Since most mineral nutrients, photosynthetic products and other biomolecules are charged molecules, water serves as the nutrient carrier and also the solvent for most biochemical reactions. Water has high heat content and high heat of vaporization, which also makes water an ideal substance for temperature regulation such as cooling through transpiration. Water is of particular importance to land plants in two other aspects. First, turgor pressure exerted against cell walls helps plants maintain their form and facilitates cell expansion and growth. Second, water is an essential reactant in photosynthesis, the principal mechanism of biomass accumulation for plants [17]. Because water molecules are highly integrated into plant physiology, water limiting conditions, normally defined as drought, are one of the major abiotic stresses that limit crop productivity [18].

In natural environments, drought stress is often accompanied and complicated by a number of other abiotic stresses, such as high temperatures, cold or salinity. High temperatures increase the driving force for transpiration while cold and soil salinity limit soil water availability to roots. To prevent water loss, land plants have evolved vapor resistant cuticles on their aerial surfaces and use stomatal apertures to regulate transpirational water loss and CO₂ uptake [19, 20].
A stomatal complex consists of a pair of guard cells surrounding a microscopic pore called a stoma or stomate; in some species, stomatal complexes also include neighboring subsidiary cells [21]. When ample water is present in the environment, stomatal pores can open in response to light to facilitate CO$_2$ uptake and photosynthesis. When water is limited, stomata close to prevent further water loss and to maintain leaf water potential. Stomatal pores are regulated by swelling or shrinking of guard cells through changing cellular water content, which in turn is driven by changes in cellular concentrations of osmotically active solutes such as K$^+$, Cl$^-$ and malate$^{2-}$ [22].

Guard cells respond to many different drought signals, including the best characterized drought hormone, abscisic acid (ABA). Molecular details of drought and ABA response in stomata are better characterized in the model species Arabidopsis as compared to drought responses in others tissues and organs in Arabidopsis or in other species [23]. This chapter describes recent advances in our understanding of genetic regulators of stomatal drought responses, with a focus on molecular mediators identified in Arabidopsis (see Table 2.1). We will also discuss recently identified genetic regulators of stomatal development in Arabidopsis. Other aspects of plant drought responses are reviewed elsewhere in this book and are not the focus of this chapter.
Table 2.1: Mutants with both stomatal phenotypes and increased drought tolerance

Among all genes mentioned in this chapter, not all have published results from whole plant drought response experiments. This table lists all published mutants for which both increased whole plant drought tolerance phenotypes and altered stomatal phenotypes have been reported. Among these 16 mutants, eight are 35S driven overexpression mutants, four are T-DNA single gene knockouts and one is a T-DNA double mutant. In response to light, five mutants have intrinsically smaller stomatal apertures, and three mutants have a smaller fraction of open stomata as compared to wild type. Nine mutants exhibit hypersensitivity in stomatal ABA responses. One mutant (MRP5) is insensitive for ABA-induced stomatal closure. All mutants tested for excised leaf water loss showed reduced water loss.

<table>
<thead>
<tr>
<th>Gene Name</th>
<th>AGI Locus Identifier</th>
<th>Mutant Type</th>
<th>Stomatal Phenotype Light</th>
<th>ABA Promotion of closure</th>
<th>Inhibition of opening</th>
<th>Leaf phenotype of Excised water loss</th>
<th>Whole plant Phenotype</th>
<th>Rewater survival rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>MYB44</td>
<td>At5g67300</td>
<td>OE</td>
<td>Smaller</td>
<td>Hypersensitive</td>
<td>NT</td>
<td>Less</td>
<td>NT</td>
<td>Better</td>
</tr>
<tr>
<td>PI5P</td>
<td>Human Gene</td>
<td>OE</td>
<td>WT</td>
<td>Hypersensitive</td>
<td>Hyposensitive</td>
<td>Less</td>
<td>Less</td>
<td>NT</td>
</tr>
<tr>
<td>CPK4</td>
<td>At4g09570</td>
<td>OE</td>
<td>NT</td>
<td>Hypersensitive</td>
<td>NT</td>
<td>Less</td>
<td>Less</td>
<td>Better</td>
</tr>
<tr>
<td>CPK11</td>
<td>At1g35670</td>
<td>OE</td>
<td>NT</td>
<td>Hypersensitive</td>
<td>Hyposensitive</td>
<td>Less</td>
<td>Less</td>
<td>Better</td>
</tr>
<tr>
<td>ABF3</td>
<td>At4g34000</td>
<td>OE</td>
<td>RN</td>
<td>NT</td>
<td>NT</td>
<td>Less</td>
<td>NT</td>
<td>Better</td>
</tr>
<tr>
<td>ABF4</td>
<td>At5g19920</td>
<td>OE</td>
<td>RN</td>
<td>NT</td>
<td>NT</td>
<td>Less</td>
<td>NT</td>
<td>Better</td>
</tr>
<tr>
<td>SDIR1</td>
<td>At5g55530</td>
<td>OE</td>
<td>RN</td>
<td>NT</td>
<td>NT</td>
<td>Less</td>
<td>NT</td>
<td>Better</td>
</tr>
<tr>
<td>MEK1</td>
<td>At4g26070</td>
<td>OE</td>
<td>Small</td>
<td>WT</td>
<td>NT</td>
<td>NT</td>
<td>NT</td>
<td>Better 3</td>
</tr>
<tr>
<td>AtMRP5</td>
<td>At1g04120</td>
<td>T-DNA KO</td>
<td>Smaller</td>
<td>Hyposensitive</td>
<td>NT</td>
<td>Less</td>
<td>Less</td>
<td>NT</td>
</tr>
<tr>
<td>CIPK23</td>
<td>At1g30270</td>
<td>T-DNA KO</td>
<td>NT</td>
<td>Hypersensitive</td>
<td>NT</td>
<td>Less</td>
<td>NT</td>
<td>Better</td>
</tr>
<tr>
<td>ABH1</td>
<td>At2g13540</td>
<td>T-DNA KO</td>
<td>NT</td>
<td>Hypersensitive</td>
<td>NT</td>
<td>Less</td>
<td>NT</td>
<td>NT</td>
</tr>
<tr>
<td>GCR1</td>
<td>At1g48270</td>
<td>T-DNA KO</td>
<td>NT</td>
<td>Hypersensitive</td>
<td>NT</td>
<td>Less</td>
<td>Less</td>
<td>Better</td>
</tr>
<tr>
<td>CBL1/CBL9</td>
<td>At4g17615</td>
<td>T-DNA KO</td>
<td>WT</td>
<td>Hypersensitive</td>
<td>NT</td>
<td>Less</td>
<td>Less</td>
<td>NT</td>
</tr>
<tr>
<td>ERA1</td>
<td>At5g47100</td>
<td>1</td>
<td>Smaller</td>
<td>Hypersensitive</td>
<td>NT</td>
<td>Less</td>
<td>NT</td>
<td>NT</td>
</tr>
<tr>
<td>AtCHL1</td>
<td>At1g12110</td>
<td>2</td>
<td>Smaller</td>
<td>WT</td>
<td>NT</td>
<td>Less</td>
<td>Less</td>
<td>NT</td>
</tr>
</tbody>
</table>

NT: Not tested in published results
OE: Over type response
RN: Reduced number of opened stomata
1. Fast neutron-generated mutant
2. γ-irradiation generated mutant
3. Better survival rate without rewatering
2.2. Arabidopsis as a model system.

Plant reactions to drought stress have been studied in many experimental species including crop species, trees, desiccation-resurrection plants and many plant model species. In recent years, the model species *Arabidopsis* has become the most widely used system to dissect the genetic basis of plant drought tolerance [23]. *Arabidopsis* has the advantage of a short life cycle and small stature. Also, *Arabidopsis* is easy to transform as compared to many other experimental and crop species. These features have allowed the creation and maintenance of a large collections of mutants [3], which are valuable tools for genetic analyses of stomatal drought response.

The genome of *Arabidopsis* has been fully sequenced and is the best annotated plant genome to date [1, 24]. This unparalleled richness of genomic information underlies many research technologies. Because stomata can only be observed under a microscope, forward-genetic screens based on stomatal physiology traits have proven somewhat difficult. In fact, many genetic regulators of stomatal functions were first identified by forward genetic screens on more easily scored traits, such as ABA sensitivity in seed germination, rather than by scoring stomatal ABA responses. In reverse genetic approaches, gene functions are first predicted by sequence homology to known genes and then tested by phenotyping plants harboring mutations in the target gene. Alternatively, many cell physiological processes involved in stomatal drought responses were first characterized in species with large stomata by chemical inhibitor analyses, and then the regulatory genes were identified in *Arabidopsis* by reverse genetic approaches [25-28]. Further, the well-annotated *Arabidopsis* genome provides information regarding gene family sizes and sequence similarities between gene family members, hence double mutants can be generated based on the evolutionary relationships between family members and used to evaluate
functional redundancy [29]. A fully sequenced genome also helps in the application of microarray [30] and newly developed sequencing-by-synthesis technologies [31]. Because naturally occurring Arabidopsis variants have different degrees of drought tolerance [32], both microarray and sequencing technologies provide opportunities to find novel genes and polymorphisms which are correlated with drought tolerance traits [33, 34]. The rich information generated by these technologies has already been used to identify new molecular mediators of drought tolerance and to correlate drought tolerance phenotypes with novel genetic loci at the whole plant level [35, 36].

2.3. How do stomates sense drought stress

Drought causes decreases in soil water availability and, in some geographic regions, reduced air relative humidity, both of which are known to reduce stomatal conductance in multiple crop species [17]. Given their location in aerial surfaces of plants, stomata cannot directly sense water limiting conditions in the soil. Instead, stomata close in response to chemical signals such as ABA. ABA is known to be synthesized in roots and transported in xylem sap to leaves when roots sense drought in the soil [37]. Ample evidence suggests that stomata can also directly respond to the changes in leaf water status, and such responses may correlate with the accumulation of ABA [38]. In addition, stomatal humidity sensing is also connected to the ABA pathway by several recently published reports [39, 40]. Because of the importance of ABA as a drought signal, a number of genetic regulators of ABA biosynthesis have been identified and they will be discussed in this section. Cytokinins, ethylene and jasmonates are known to be drought signals that act on guard cells, although none of these hormones has been found to override the dominant role of ABA in stomatal drought response [41, 42]. Molecular components mediating stomatal response to other hormones and interactions between ABA and other hormones during drought will also be discussed (see Figure 2.1).
Figure 2.1. Stomatal drought response is mediated by plant hormones. Solid lines indicate regulation of stomatal opening and closure. Dashed lines indicate regulation of stomatal development. See text for details.

2.3.1 ABA as a drought hormone

ABA is a 15 carbon (C15) molecule derived from the C40 carotenoid precursor zeaxanthin. The Arabidopsis gene \textit{ABA1} is a zeaxanthin epoxidase which catalyzes the conversion of zeaxanthin to violaxanthin [43, 44]. Both drought and ABA induce \textit{ABA1} gene expression. Violaxanthin is converted to neoxanthin and then 9’-cis-neoxanthin and 9’-cis-violaxanthin, which are both cis-xanthophylls (C40). \textit{ABA4} was identified by map based cloning as a positive regulator of neoxanthin synthase [45]. Unlike \textit{ABA1}, \textit{ABA4} is not induced by drought, but knockout of either of these two genes eliminates dehydration-induced ABA accumulation [45]. Consistent with these results, both \textit{aba1} and \textit{aba4} mutants showed lower leaf temperatures than wild type after 3 days of drought stress [45].
The ABA specific metabolic pathway starts with the cleavage of cis-xanthophylls into C15 xanthoxin and a C25 byproduct catalyzed by 9-cis-epoxycarotenoid dioxygenases (NCEDs) [44]. In Arabidopsis, tissue specificity of ABA biosynthesis emerges at this step, since a family of nine NCED genes has been identified in the genome [46]. Among these nine genes, five NCEDs have been studied by promoter-GUS assay [47]. AtNCED2 is expressed in guard cells from senescing leaves, while AtNCED3 is expressed in guard cells in cotyledons, hypocotyls, and petioles. Both AtNCED2 and AtNCED3 are expressed in roots, implying that both of these genes play roles in root ABA biosynthetic processes. Surprisingly, none of the five AtNCED genes was found by reporter gene analysis to be expressed in guard cells from rosette leaves or cauline leaves.

ABA2 and AAO3 are enzymes downstream of AtNCEDs, and catalyze the transformation of xanthoxin to abscisic aldehyde and abscisic aldehyde to abscisic acid, respectively [48, 49]. AAO3 activity also requires another gene, ABA3, which encodes a MoCo sulfurase, and the homolog of this gene, flacca, was one of the first ABA biosynthetic genes found by genetic approaches in tomato, where its mutation results in a wilty phenotype [50]. Among all ABA biosynthesis genes in Arabidopsis, only NCED3 was strongly induced by drought in leaf vascular tissue [46, 51]. The expression of AAO3 was detected in guard cells and significantly induced in guard cells upon drought treatment, consistent with a previous report that guard cells can themselves synthesis ABA [52].

ABA catabolism includes both conjugation and oxidation into inactive forms. The predominant pathway of ABA oxidation is thought to be 8'-hydroxylation. This process is catalyzed by a family of four CYP707A genes, which can be induced by seedling dehydration [53]. CYP707A1 and CYP707A3, but not CYP707A2, were shown to be induced in guard cells and vascular tissue, respectively, by high humidity. A cyp707a1 cyp707a3 double mutant showed stronger defects in stomatal opening responses to high humidity than either cyp707a1 or
cypher 707a3 single mutants, suggesting that these two genes have synergistic functions [54].

Because high air humidity could be the first sign of drought release due to precipitation, these results highlight the role of ABA oxidation in drought release responses of guard cells.

ABA remobilization from conjugated forms seems to be involved in stomatal regulation. In *Arabidopsis*, ABA was shown to be released from a biologically inactivated, glucose conjugated form by a β-glucosidase, BG1 [55]. Dehydration, and even moderate changes in air humidity, can rapidly activate BG1 by inducing BG1 protein polymerization. Under darkness, T-DNA insertion mutant bg1 showed defective stomatal closure, which could be rescued by exogenous ABA [55]. The bg1 mutant loses more water in detached rosettes and is more sensitive to drought stress than wild type [55].

Further experiments on the temporal expression patterns of genes involved in ABA biosynthesis and catabolism, their responses to different levels of drought stresses, their response to drought release, and effects of ABA metabolism on stomatal movement will provide valuable insight regarding the interconnection of ABA metabolic pathways and stomatal drought responses.

### 2.3.2 Other hormones in stomatal drought responses

Besides ABA, cytokinins are also well accepted root-to-shoot signaling hormones which participate in plant drought response [56]. In multiple species, drought stresses decrease cytokinin concentrations in xylem sap, while high concentrations of cytokinins antagonize ABA effects on stomatal closure and can induce stomatal opening [57]. In *Arabidopsis*, three cytokinin receptor histidine kinases (CRHK) mutants, *ahk2*, *ahk3*, and *cre1*, are hypersensitive to ABA and osmotic stress and thus more tolerant to drought stresses [58]. These CRHKs also showed functions antagonistic to AHK1, an osmotic sensing receptor histidine kinase [58]. Both of these experiments were carried out at the whole plant level, therefore, whether any of these cytokinin
receptors directly function in guard cells remains unknown. Cytokinins inhibit ABA-induced stomatal closure in wild-type but not in the ethylene insensitive mutant ein3-1. In the cytokinin over-producing mutant, amp1-1, ABA-induced stomatal closure is impaired but can be restored through inhibition of ethylene biosynthesis [59]. While these results suggest that cytokinins inhibit ABA-induced stomatal closure through ethylene biosynthesis, the possibility that cytokinin receptor kinases are also activated in guard cells cannot be ruled out.

The role of ethylene in plant drought response is unclear. In different species, ethylene production differentially depends on the duration and intensity of drought [42]. In Arabidopsis, ethylene inhibits ABA-induced stomatal closure [60], and also mediates an antagonistic effect of cytokinin on ABA-induced stomatal closure as described above [59]. However, ethylene alone was shown to induce stomatal closure in Arabidopsis, and this process is mediated by H$_2$O$_2$ [61]. Ethylene receptor histidine kinase ETR1 and signaling proteins EIN2 and ARR2 are involved in ethylene induced stomatal closure, and AtrbohF, a NADPH oxidase, seems to be an intermediate protein in ABA and ethylene crosstalk in guard cells [62]. Another histidine kinase, AHK5, is downstream of ethylene-induced ROS production but not ABA-induced stomatal closure [63]. In addition to regulating stomatal aperture, ethylene treated plants or ethylene over-production mutants showed increased stomatal density in Arabidopsis [64].

Jasmonic acid and methyl jasmonate (MeJA), collectively called jasmonates (JA), increase rapidly in response to drought [65]. Unlike drought-induced ABA production, plant JA content drops after an initial peak, which suggests a transient role of JA in plant drought response [65-67]. MeJA can induce stomatal closure in Arabidopsis, but such effects may depend on the plant growth conditions [68]. MeJA-induced stomatal closure is mediated by a number of common intracellular components of stomatal ABA responses, but is also mediated by several MeJA specific components. In terms of common mediators, both MeJA and ABA induce stomatal closure by inducing ROS and NO production, which in turn activate membrane Ca$^{2+}$
permeable channels, anion channels and outward K$^+$ channels in guard cells [69, 70]. Also, the induction of ROS in guard cells by either MeJA or ABA is both dependent on protein phosphatase RCN1 and AtrbohD/F [71]. However, MeJA but not ABA regulation of ROS, NO and ion channel activities is impaired in two JA insensitive mutant, jar1 and coi1 [70], while ABA but not JA induced ROS and NO production is mediated by protein kinase OST1 [72].

2.4. Signaling events inside guard cells in response to drought.

Drought stress is perceived by guard cells as a composite signal from multiple plant hormones but predominantly from ABA, which activates a wide range of intracellular messengers such as calcium, nitric oxide (NO), reactive oxygen species (ROS) phosphatidic acid (PA) and intracellular pH changes. A number of genetic regulators upstream of these intracellular messengers have been identified in *Arabidopsis*. In addition, many of these intracellular messengers can modulate membrane permeability through direct or indirect regulation of plasma membrane and vacuolar membrane ion channels, transporters and pumps. Some of these intracellular messengers, such as calcium and proton concentrations, are themselves regulated by membrane channels and pumps. Membrane channels and transporters mediate the uptake and release of osmotically active solutes such as K$^+$, Cl$^-$ and malate$^{2-}$ from guard cell cytosol and vacuoles, resulting in changes of cellular water content and guard cell volume, and eventually leading to changes in stomatal apertures (see Figure 2.2).
Figure 2.2. Cellular mediators of stomatal drought response.
The ABA-mediated drought stimulus acts on signaling proteins (see section 2.5.1.2 and 2.5.3) and membrane transport related proteins (see section 2.4.1). Signaling proteins regulate membrane transport and intracellular enzymes. Although in mammalian systems, membrane signaling proteins can directly activate intracellular protein kinases, this type of connection has not yet been found in guard cells. Changes in secondary messengers are regulated by enzymes that catalyze the formation of messengers or by membrane transport (see section 2.4.2 to 2.4.4). Secondary messengers regulate many downstream signaling components. Protein kinases and phosphatases are also involved in stomatal drought response (see section 2.5.1 and 2.5.2). Protein kinases and phosphatases can regulate membrane transporters, which is omitted from this figure for clarity. Transcription factors regulate both drought induced gene expression (see section 2.6.1) and stomatal development (see section 2.6.2). mRNA binding proteins are involved in stomatal ABA and humidity responses (see section 2.5.5).

HT G proteins: Heterotrimeric G proteins
Rboh: NADPH oxidases
NR: Nitric reductases
InsPs: Inositol phosphates
PIPs: Phosphatidylinositol phosphates
CBP: mRNA cap binding proteins
2.4.1 Ion channels and other membrane transport related proteins.

Light activation of membrane H⁺ ATPases causes hyperpolarization of the plasma membrane, which is the driving force for ion flow into guard cells. Two dominant mutants ost2-1D and ost2-2D, which each harbor mutations in the guard cell-expressed H⁺ ATPase gene, AHA1, were found to have reduced leaf surface temperatures by a thermal imaging based screen. Stomata of ost2 mutants cannot close in response to ABA, due to constitutively active H⁺ ATPase activity [73]. This result suggests that down regulation of H⁺ ATPase activity is an important step in stomatal drought response [74].

KAT1 is a well-characterized inward K⁺ channel subunit highly enriched in guard cells [75], which forms functional multi-protein complexes with KAT2 [76] and potentially several other K⁺ channel proteins in guard cells [77, 78]. At least five K⁺ channel proteins are expressed in guard cells [77], and knockout of KAT1 does not impair stomatal opening. However, plants that overexpress dominant negative alleles of KAT1 or KAT2 have reduced inward K⁺ current and reduced water loss [79, 80], suggesting that K⁺ influx mediated by inward K⁺ channels is important for light-induced stomatal opening and plant drought response. ABA regulates inward K⁺ channel activity not only through secondary messengers, but also through vesicle trafficking to and from the cell membrane [81]. Two syntaxin proteins have been studied in detail in Arabidopsis. SYP121 was found to be directly involved in KAT1 membrane trafficking [82], while SYP61 is enriched in guard cells, and is important for osmotic stress response [83].

In Arabidopsis, GORK is the major outward K⁺ channel gene expressed in guard cells. T-DNA knockout and dominant negative mutants of GORK showed impaired outwardly rectifying K⁺ channel activity [84]. gork mutants also have increased water loss in both excised leaf and whole plant experiments [84]. Besides plasma membrane K⁺ channels, plants also have ion
channels localized in internal membranes, including AtTPK1, which is a two-pore potassium channel mainly located in the tonoplast. Knockouts of AtTPK1 have reduced stomatal closure rate as compared to wildtype, which is consistent with AtTPK1 mediated K⁺ release from vacuole to cytoplasm [85].

Besides K⁺ channels, anion channels are also important for stomatal movement in response to environmental stimuli. SLOW ANION CHANNEL-ASSOCIATED 1 (SLAC1) was recently cloned and proposed to encode a guard cell anion channel [26, 27]. Knockouts of SLAC1 showed impaired stomatal responses to a range of environmental signals including ABA and humidity. Although knockouts of SLAC1 showed abolished anion channel activity in planta, heterologous expression of SLAC1 in yeast cannot complement a yeast malate uptake deficient mutant, and expression of SLAC1 in Xenopus oocytes does not result in anion channel activity [26, 27]; one explanation is that additional components are required for SLAC1 anion channel activity. ATP-binding cassette proteins have been found to modulate guard cell anion channel activity and thus are potential partners for SLAC1 [86]. Two ATP-binding cassette (ABC) transporters, AtMRP4 and AtMRP5, have been studied in detail in Arabidopsis. Mutants of AtMRP4 and AtMRP5 have opposite phenotypes. In atmrp4 mutants, stomatal apertures are larger than wild type under both light and darkness while stomatal ABA response is similar to wild type, therefore plants loss more water under drought conditions [87]. In contrast, atmrp5 mutants showed reduced water loss [88]. Although ABA-induced stomatal closure is impaired in atmrp5 mutants, light-induced stomatal opening is also reduced, which could be the main reason for the reduced water loss observed at the whole plant level [88].

Besides the well-known K⁺ channels and anion channels, a number of other relevant membrane transporters have also been identified. For example, AtCHX20, a cation/H⁺ exchanger, was found to be a positive regulator of light-induced stomatal opening [89]. Knockouts of a nitrate transporter, AtCHL1/NRT1, implicated in NO₃⁻ uptake during stomatal opening, have
enhanced drought tolerance [90]. In addition, a PLEIOTROPIC DRUG RESISTANCE 3 transporter, AtPDR3 [91], and a sucrose transporter, AtSUC3 [92], are both highly expressed in guard cells, although their roles in guard cell physiology remain unknown.

2.4.2 Calcium

ABA induces cytosolic Ca\(^{2+}\) increases through the activation of Ca\(^{2+}\) permeable channels in the plasma membrane and through the activation of Ca\(^{2+}\) release from internal stores. Elevated cytosolic Ca\(^{2+}\) ions can inhibit inward K\(^+\) channels and H\(^+\) pumps at the plasma membrane while activating anion channels [22].

*Arabidopsis* CYCLIC NUCLEOTIDE GATED CHANNEL2 (CNGC2/DND1) mediates cAMP-induced Ca\(^{2+}\) influxes through plasma membrane channels [93]. Despite the importance of Ca\(^{2+}\) channels in stomatal ABA response, guard cell phenotypes have not yet been tested under ABA or drought treatment in the *dnd1* mutant. Ca\(^{2+}\) release from internal stores may be partially mediated by a tonoplast channel, the two pore channel TPC1 [94-96]. In addition, a thylakoid-localized calcium sensor (CAS1) may participate in Ca\(^{2+}\)-induced-Ca\(^{2+}\)-release from chloroplasts [97, 98]. Both the *cas1* mutant and the *tpc1* mutant showed impaired extracellular Ca\(^{2+}\) induced stomatal closure, but these two mutants did not have impaired stomatal ABA response [95, 97]. Little is known about the importance of Ca\(^{2+}\)-induced closure under physiological conditions, thus the role of these two proteins in plant drought response is not clear [96, 98].
2.4.3 ROS and NO

ABA activates plasma membrane Ca\(^{2+}\) channels by inducing cytosolic H\(_2\)O\(_2\) production by the NADPH oxidases, AtrbohD and AtrbohF in *Arabidopsis* guard cells [99-101]. AtrbohD/F mediated ROS production is also involved in stomatal ethylene [62] and jasmonate signaling [70]. Other evidence also suggests that a small GTPase ROP2 and phosphatidic acid (PA) pathway is involved in ABA-induced ROS production [102]. ROS also inhibit the enzymatic activities of other mediators of stomatal ABA responses, such as plasma membrane H\(^+\) ATPase activation [103], and protein phosphatases ABI1 and ABI2 [104]. Further, an *Arabidopsis* glutathione peroxidase, AtGPX3, interacts with ABI1 and ABI2 and acts as a positive regulator of stomatal ROS responses [105].

In *Arabidopsis*, ABA-induced NO production in guard cells is mainly mediated by NIA1 and NIA2 [61], and also requires H\(_2\)O\(_2\) production from the AtrbohD/F pathway [106]. NO activates anion channels and inhibits inward K\(^+\) channels through cADPR- and cGMP-dependent intracellular calcium release [107, 108], whereas NO inhibition of outward K\(^+\) channels may be through direct protein modification [109].

2.4.4. Other small intracellular molecules.

ABA induced cytosolic Ca\(^{2+}\) oscillation and stomatal closure is partially mediated by phospholipase C (PLC), which hydrolyses phosphatidylinositol-4,5-bisphosphate (PIP2) into inositol-1,4,5-trisphosphate (InsP3) (which induces Ca\(^{2+}\) release) and diacylglycerol [110, 111]. Inositol polyphosphate 5-phosphatase (InsP 5-ptase) catalyzes the hydrolysis of InsP3 and terminates the ABA induced InsP3 signaling [112]. Transgenic plants with overexpression of InsP 5-ptase showed better whole plant drought tolerance and less water loss from excised leaves.
than wild type. Transgenic plants are hypersensitive to ABA-induced stomatal closure but insensitive to ABA inhibition of stomatal opening [112]. These results suggest that, under this particular experimental condition, hypersensitivity to ABA-induced stomatal closure is more important for the observed whole plant drought tolerance phenotype.

In a functional proteomic study, TGG1 was identified as one of the most abundant proteins in guard cells [68]. TGG1 encodes a myrosinase, which cleaves glucosinolates to produce toxic compounds such as isothiocyanates. T-DNA knockouts of tgg1 are hyposensitive to ABA-inhibited stomatal opening, consistent with a lack of ABA inhibition of inward K+ channels in guard cells of tgg1 mutants [68]. The glucosinolate-myrosinase system is well known for its function in plant biotic stress response [113]. Because of the dual role of TGG1 in both biotic and abiotic stress responses, TGG1 is a good candidate for engineering pest and drought resistant crop plants.

2.5. Cell signaling mutants with altered stomatal responses.

Stomatal drought response involves not only ion channels and secondary messengers, but also a large number of cell signaling proteins including protein kinases, protein phosphatases, G-proteins, and farnesyl transferase. Interestingly, many of these signaling proteins are also related to stomatal humidity response, which is also discussed in this section (see Figure 2.2).

2.5.1 Protein Kinases.

2.5.1.1 SnRK protein kinases.

Among many different types of protein kinases, members of the SNF1-related protein kinase (SnRK) family are conserved mediators of ABA and osmotic stress responses in different
plant species. Proteins from this family, including *Vicia faba* AAPK, barley PKABA, soybean SPK1/2, tobacco OSAK, rice SAPK, and those from *Arabidopsis* [114-119] are activated by ABA, osmotic stress or both in various plant tissues. Among these genes, *OST1* and *AAPK* are highly expressed in guard cells and play roles in stomatal function. AAPK was first identified as an ABA-activated protein kinase by in-gel kinase assay followed by mass spectrometry based protein identification, while its ortholog OST1 was isolated in *Arabidopsis* by a thermal imaging aided forward genetic screen. Originally, two mutant alleles of *OST1*, *ost1-1* and *ost1-2*, were isolated by map based cloning. Both mutants contain point mutations, and the mutant plants lack stomatal ABA responses but not stomatal light and CO2 responses [117]. OST1 protein is activated by ABA in guard cells and roots in a calcium independent manner. This activation is blocked by the *abi1-1* mutation but not by the *abi2-1* mutation [120]. Also, ABA-induced ROS production in guard cells is interrupted by *ost1* mutation, while direct application of ROS and calcium can cause stomatal closure in *ost1* mutants. This evidence indicates that OST1 is upstream of ROS production and downstream of *abi1-1* [117]. Subsequent experiments showed that the ABI1 protein physically interacts with the C-terminal domain II of OST1, which is required for OST1 function in guard cell ABA responses. On the other hand, the OST1 C-terminal domain I is not required for ABA response but is required for osmotic activation of OST1 kinase activity [120]. OST1 and several other SnRK2-like proteins phosphorylate AREB1, a transcription factor that binds to ABA response elements (ABRE). These results support a critical role of the OST1 kinase in guard cell signaling.

2.5.1.2 LRR receptor kinases

Receptor-like kinases (RLKs) are transmembrane kinases that function in cell-to-cell communication and environmental signal perception in many eucaryotes [121]. The *Arabidopsis* genome contains more than 400 RLKs, with 216 Leucine-rich repeat (LRR) containing receptor kinases representing the largest RLK family. In general, LRR receptor kinases have three
domains: one LRR containing extracellular domain, one intracellular serine/threonine (ser/thr) protein kinase domain, and one single-pass transmembrane domain [121].

*Arabidopsis* RPK1 is a LRR receptor kinase that mediates ABA responses in plant germination, growth and stomatal responses. The transcript level and protein level of RPK1 are both upregulated by ABA. In two independent mutant alleles of *RPK1*, ABA-induced stomatal closure was impaired [122]. However, components downstream of RPK1 and the effects of rpk1 mutation on plants under drought conditions remain unknown.

The *Arabidopsis* ERECTA (*ER*) gene is a well known LRR-receptor kinase that controls plant lateral organ size and flower development by regulating plant cell proliferation [123]. By quantitative trait loci (QTL) analysis, the *Arabidopsis ERECTA* gene was found to be a genetic regulator of transpirational efficiency (TE) [124], which is a measure of plant water use efficiency. All three *erecta* mutants used in the study showed increased stomatal conductance, partly due to increased number of stomata on the leaf surface. *erecta* mutants also have fewer mesophyll cells and reduced photosynthetic capacity as compared to plants with a functional *ERECTA* gene. Both increased stomatal conductance and decreased photosynthetic capacity negatively affect TE, consequently, erecta mutants have lower TE in both well watered and drought conditions [124]. Because the ABA sensitivities of erecta mutants have not yet been tested, to date, evidence suggests that *ERECTA* affects plant TE through multiple morphological changes.

### 2.5.1.3 CDPK and CIPK network

Cytosolic Ca$^{2+}$ is one of the key secondary messengers for multiple cellular responses in guard cells. Ca$^{2+}$ dependent protein kinases (CDPK) [125], and calcineurin B-like (CBL) proteins [126] are important intracellular calcium sensors, propagating the cytosolic Ca$^{2+}$ signal to downstream targets such as membrane ion channels and transcription factors. The Arabidopsis genome encodes 34 CDPKs, each of which contains a kinase domain, an autoinhibitory domain
and a calmodulin-like domain [125]. The calmodulin-like domain binds Ca\(^{2+}\) when cytosolic free Ca\(^{2+}\) concentration is elevated, thus causing a conformational change that releases the autoinhibitory domain from the kinase domain and activates the CDPK.

Recently, four *Arabidopsis* CDPKs were found to be related to stomatal functions. CPK3 and CPK6 were first identified by PCR from a guard cell-enriched cDNA library, and their expressions in guard cells were later confirmed by both RT-PCR and microarray analysis [29]. In both single mutants, *cpk3* and *cpk6*, and double mutants *cpk3cpk6*, Ca\(^{2+}\) activation of slow-anion channels was impaired. Furthermore, in *cpk3cpk6* double mutants, ABA activation of S-type anion channel is strongly inhibited. In guard cells, one way of increasing cytosolic Ca\(^{2+}\) concentration is through activation of plasma membrane Ca\(^{2+}\) permeable-channels, which can be activated by ROS. In the *cpk3cpk6* double mutant, ABA but not ROS activation of Ca\(^{2+}\) channels was inhibited. Finally, in both single mutants and the double mutant, ABA-induced stomatal closure was impaired [29].

The functions of two other CDPKs, CPK4 and CPK11, have been characterized for various ABA responses in different organs and tissues [127]. For drought related functions, the single mutants and double mutant of these two CDPKs are less sensitive to ABA-induced stomatal closure as compared to wild type, while overexpression of either CDPK conveys ABA hypersensitivity. In water loss experiments, single and double mutants lost more water, while overexpression lines conserved more water than wild-type plants. Furthermore, in drought-rewater experiments, overexpression lines had higher survival rates than wild-type plants whereas all single and double mutants died under the same condition. Two ABA response transcription factors, ABF1 and ABF4 can be phosphorylated by these two CDPKs, which connects these CDPKs’ functions to transcriptional regulation [127].

Unlike the CDPKs, which can be directly activated by Ca\(^{2+}\), CBL proteins do not contain kinases domains and cannot directly phosphorylate downstream effectors. Instead, after Ca\(^{2+}\)
binding at EF-hand domains, CBL proteins can bind and activate CBL-interacting protein kinases (CIPKs). The Arabidopsis genome encodes 10 CBL proteins and 25 CIPKs [128], which form an interconnected network, with multiple CBLs interacting with the same CIPK. T-DNA insertional mutants for all 25 CIPKs in Arabidopsis were screened for their phenotypes under drought conditions [129]. Among the CIPK mutants tested, the cipk23 mutant had the best survival rate (70%) while wild-type had only a 20% survival rate under the same experimental condition. A number of drought or ABA inducible genes were induced to a similar level in wild type and cipk23 mutants, and ABA content was not elevated in the cipk23 mutant. However, cipk23 mutant stomata are hypersensitive in both ABA-induced stomatal closure experiments and ABA-inhibited stomatal opening experiments. Earlier publications had shown that the upstream factors of CIPK23 are CBL1 and CBL9 [128], both of which can interact with CIPK23 and in turn, activate a K⁺ channel, AKT1, in roots [130]. The double mutant, cbl1cbl9, but not cbl1 or cbl9 single mutants, conserved more water in water loss experiments as compared to wild type [129]. Since CIPK23, CBL1 and CBL9 coexpress in guard cells [129], these results suggest that CBL1 and CBL9 function synergistically upstream of CIPK23 in the regulation of stomatal ABA response and plant drought tolerance. In another study, CBL1 (called SCaBP5 in the publication) and CIPK15 (called PKS3) were shown to interact in yeast two hybrid experiments. RNAi lines of cbl1 or cipk15 were hypersensitive to ABA in stomatal closure, and thus lost less water as compared to wild type [131]. However, the stronger effect of CBL1 RNAi lines than knockout lines could be due to off target effects of RNAi. CIPK15 can interact with an APETALA2/EREBP-type transcription factor, AtERF7, which is a negative regulator in stomatal ABA responses [132].

2.5.1.4 MAP kinases cascade.

Mitogen-activated protein kinase (MAPK) cascades are conserved signaling modules, consisting of three types of protein kinases: MAPK, MAPK kinase (MAPKK), and MAPKK
kinase (MAPKKK). As indicated by their names, MAPKKKs phosphorylate MAPKs and MAPKs phosphorylate MAPKs, which propagate signals to downstream effectors. Plant MAPK cascades are known to be involved in multiple biotic and abiotic stress responses, hormonal responses, and plant development [133]. Recently, two members of Arabidopsis MAPK cascades, MPK3 [134] and MEK1 [135], were found to be related to stomatal ABA responses. Antisense lines of MPK3 were found to be less sensitive to ABA inhibition of stomatal opening under normal conditions, while ABA promotion of stomatal closure was reduced in these lines if cytosolic pH was also clamped. Further experiments showed that stomata in these RNAi lines are less sensitive to H$_2$O$_2$-induced stomatal closure and inhibition of opening. However, ABA-induced H$_2$O$_2$ production was not affected. These results suggest that MPK3 is involved in cytosolic pH regulation, and is downstream of H$_2$O$_2$ production [134]. Unlike MPK3, MEK1 was found to be upstream of H$_2$O$_2$ production in Arabidopsis. A mek1 T-DNA insertional mutant is less sensitive to ABA induced closure and hypersensitive to drought stress, while overexpression lines of MEK1 are more resistant to drought treatment [135]. While MPK3 and MEK1 are related to stomatal ROS response, several other members of MAPK cascades have been found to be related to stomatal development, and will be discussed in a later section.

2.5.2 Phosphatases.

Plant protein phosphatases catalyze dephosphorylation reactions. Members from three categories of protein phosphatases, protein phosphatases 2C (PP2C), protein phosphatases 2A (PP2A) and phosphotyrosine phosphatase (PTP) are involved in stomatal drought response [136, 137].

2.5.2.1 PP2C
Type 2C protein phosphatases are important modulators in stomatal ABA responses. So far, four *Arabidopsis* PP2Cs, ABI1, ABI2, HAB1 and AtPP2CA, have been studied in detail. By forward genetic screen, dominant mutants *abi1-1* and *abi2-1* were both found to be ABA insensitive in germination assays [138]. These loci were later found to encode two closely related PP2Cs, ABI1 and ABI2, whose phosphatase activities were diminished by point mutations at the same conserved amino acid in *abi1-1* and *abi2-1* [139-143]. Although phenotypes of these mutants suggest that both genes are positive regulators for various plant ABA response, experiments using recessive and loss of function mutants of *ABI1* and *ABI2* indicate that both genes are negative regulators of ABA signaling [144, 145]. Due to this difference in ABA sensitivities between different types of mutants, one should be cautious about the placement of these two proteins within the ABA signaling network. However, biochemical experiments using wildtype ABI1 and ABI2 showed that both proteins can be modulated by a number of signaling molecules downstream of ABA. For example, ABA causes increased cytosolic ROS, which inhibits the phosphatase activity of ABI1 and ABI2, while ABA also causes cytosolic pH increase, which increases the phosphatase activity of ABI1 and ABI2 [104, 146, 147]. In yeast-two-hybrid experiments, ABI2, and, to a lesser extent, ABI1, bind to protein kinase CIPK15, which is a negative regulator of stomatal ABA response [131]. Also, an important ABA-induced secondary messenger, phosphatidic acid (PA), interacts with ABI1 and represses ABI1 activity [148]. This process is related to ABA and G protein regulation of stomatal aperture and will be discussed in detail in a later section.

Genetic analysis of *abi1* and *abi2* recessive mutants also showed that these two genes have synergistic functions in ABA-induced stomatal closure and hence plant drought response [145]. Interestingly, such synergistic interaction was also observed between ABI1 and another *Arabidopsis* PP2C, HAB1 [149]. As a highly ABA-induced ABI1/2 homologue [139], HAB1 is expressed ubiquitously in all plant tissues and is highly expressed in ABA target tissues,
including guard cells, as shown by reporter gene [150] and microarray [30] analyses. Although stomata are hypersensitive to ABA promotion of closure in one hab1 knockout line [30], another knockout line shows a wild type response in water loss assays [150]. Double mutants of both hab1 and one of the abi1-2 or abi1-3 knockout lines showed stronger stomatal ABA response and better drought tolerance than single recessive mutants of either gene or wild type [149]. Such synergistic interactions between PP2C genes are likely due to their redundant yet not completely overlapping functions in plant ABA and drought response.

The last PP2C protein discussed in this chapter is AtPP2CA, which interacts with K+ channels AKT2 and AKT3, and thus may be involved in regulation of plant K+ homeostasis [151, 152]. atpp2ca knockout plants are hypersensitive to ABA promotion of stomatal closure, but perform similarly to wildtype plants in water loss experiments. Overexpression of AtPP2CA causes stomatal ABA insensitivity and increased plant water loss [153]. Interestingly, plants with overexpression of HAB1 also showed ABA insensitivity while knockout lines of the same gene were not significantly different from wildtype. These results again suggest redundant functions between PP2C genes. Finally, a mutant with reduced AtPP2CA phosphatase activity also has increased ABA accumulation in seeds [154]. Whether mutants of other PP2C genes also affect ABA accumulation is not known.

2.5.2.2 PP2A

RCN1 is a guard cell-enriched protein phosphatase 2A which is also expressed in other tissues including mesophyll cells, as revealed by promoter GUS analysis [28]. In guard cells, an rcn1 T-DNA knockout impairs two important intermediate ABA responses, cellular ROS increase and intracellular Ca$^{2+}$ increase; loss of these responses in turn causes impaired anion channel and inward K$^+$ channel responses to ABA. Thus, rcn1 mutant guard cells are less sensitive to ABA-induced stomatal closure [28]. MJ-induced stomatal closure is also impaired in the rcn1 mutant
and MJ no longer promotes ROS production, which suggests ABA and MJ crosstalk in guard cells is mediated by a RCN1-ROS pathway [71].

2.5.2.3 PTP

Experiments using a protein tyrosine phosphatase (PTP) inhibitor, phenylarsine oxide, supported a role of PTP in stomatal ABA response [155]. In a recent study, a T-DNA insertional mutant, \textit{phs1-3}, which has reduced expression of tyrosine phosphatase \textit{PHSI}, was found to have closed stomata under light when the measurement was taken immediately after harvest. When stomata of \textit{phs1-3} were opened using an opening medium under light, they showed hypersensitivity in ABA inhibition of light-induced stomatal opening [156]. Further investigation using knockout lines of \textit{PHSI} may provide more insight into the function of PTP in stomatal ABA response.

2.5.3 G proteins

Several plant G-proteins, including members of the heterotrimeric G-protein complex and three small GTPases, have been found to mediate stomatal drought response. Heterotrimeric G-proteins are comprised of three subunits, Ga, Gβ and Gγ, whereas proteins in the small GTPase family are monomeric proteins [157].

2.5.3.1 Heterotrimeric G proteins

Canonical heterotrimeric G-protein complexes are signaling mediators that transduce extracellular signals perceived by G protein coupled receptors (GPCR) to intracellular effectors such as phospholipases and ion channels. The GDP-bound Ga subunit, Gβ subunit and Gγ subunit form a heterotrimer, which binds to cytosolic regions of the GPCR. Activation of the GPCR causes the exchange of GDP for GTP on the Ga subunit, and also the dissociation of GTP-bound Ga from the Gβγ dimer. Both the Ga-GTP and the Gβγ dimer can activate downstream
effectors until the intrinsic GTPase activity of Ga hydrolyzes the GTP into GDP. Then, inactivated Ga recruits Gβγ back into the inactive heterotrimeric complex [157, 158].

The Arabidopsis genome encodes one canonical Ga subunit, GPA1, one Gβ subunit, AGB1 and two known Gγ subunits, AGG1 and AGG2 [159-161]. T-DNA insertional mutants of GPA1 and AGB1 are both insensitive to ABA inhibition of stomatal opening, but not ABA induction of stomatal closure [25, 162]. These phenotypes are consistent with the eletrophysiological evidence that gpal and agbl mutants are both insensitive to ABA inhibition of inward K⁺ channels and conditionally insensitive to ABA activation of anion channels (i.e. insensitive under strong cytosolic pH buffer). In addition, the gpal mutant has wider stomata in normal growth condition and loses more water as compared to wild type plants. This evidence together supports a role of GPA1 and AGB1 in stomatal drought response [25, 162]. However, the knockout mutants of two Gγ subunits, AGG1 and AGG2, do not show the same stomatal phenotype as gpal and agbl in response to ABA [163]. Such observation may suggest plant heterotrimeric G-proteins function differently from mammalian G proteins or that additional Gγ subunits exist.

Sphingosine-1-phosphate (S1P) is one signaling molecule upstream of GPA1-mediated stomatal drought responses [164-166]. Sphingosine kinase (SPHK) catalyzes S1P production, and SPHK activity is enhanced by ABA [165]. S1P inhibits stomatal opening and promotes stomatal closure by inhibiting inward K⁺ channels and activating outward anion channels in wildtype plants, whereas in the gpal mutant, S1P cannot regulate ion channels or stomatal apertures [165]. In addition to S1P, GCR1 is another upstream component of GPA1-mediated stomatal signaling. GCR1 has sequence homology to one class of mammalian GCPRs, and was shown to interact with GPA1 both in vitro and in vivo. Leaves from gerl mutants lose less water as compared to wild type. Also, gerl mutants showed better tolerance to drought treatment and higher survival rate after re-watering. In contrast to the gpal mutant, stomata of gerl mutant are hypersensitive
to both ABA and S1P, which suggests that GCR1 is a negative regulator of G protein signaling in guard cells [167].

Recently, several other candidate plant GPCRs have been cloned [168]. GCR2 had been proposed as an ABA receptor that interacts with GPA1 [169]. However, the function of GCR2 is still under debate [170-172]. Besides GCR2, two novel GPCR-Type G proteins, GTG1 and GTG2 have been identified and shown to bind ABA [173]. The double mutant *gtg1 gtg2* is hyposensitive to ABA in assays of ABA induced stomatal closure, but has a wild type response in ABA inhibition of stomatal opening. GTG1 and GTG2 each can interact with GPA1 in vitro and in vivo, which supports their roles in G protein signaling [173].

Phospholipase D (PLD) activity and the resultant product, phosphatidic acid (PA) are also related to GPA1 function in stomatal ABA response. In wild type plants, ABA promotes PLD activity and PA production, which in turn promote stomatal closure and inhibit stomatal opening. Arabidopsis PLDα1 physically interacts with GPA1 and can stimulate the GTPase activity of GPA1. While GPA1 binding activates PLDα1 activity, adding GTP in the reaction negatively regulates GPA1-PLDα1 binding [174]. Stomatal ABA and PA responses were tested in two single mutants, *pldα1* and *gpa1*, and double mutant *pldα1 gpa1* [175]. In these mutants, plants harboring *gpa1* mutation are hyposensitive to ABA- and PA-inhibition of stomatal opening. Plants harboring *pldα1* mutation are hyposensitive to ABA-induced stomatal closure, and this insensitivity can be rescued by exogenous PA treatment [175]. These results suggest PA is upstream of GPA1 in ABA inhibition of stomatal opening but not in ABA induced stomatal closure. The observed stomatal phenotype of *pldα1* knockout mutants can be partially explained by the interaction between PA and ABI1, as the latter is a negative regulator of stomatal ABA response. PA interacts with ABI1 and reduces the effect of ABI1 by both inhibition of phosphatase activity and by relocation of ABI1 from cytosol to cell membrane [148]. ABA induced a similar level of stomatal closure in wild type and *abi1* knockout mutant, but not in an
abi1 mutant expressing PA insensitive ABI1R73A protein. Also, ABA failed to induce stomatal closure in the pldα1 mutant, but this phenotype was reversed in the pldα1 abi1 double mutant. All evidence above implies that ABA-PLD-PA-ABI pathway is important for ABA-induced stomatal closure. However, both abi1 knockout and ABI1R73A mutants showed wildtype ABA response in opening experiments, suggesting that ABI1 is not related to the ABA inhibition of opening pathway [175].

2.5.3.2 Small GTPase

Two small GTPases, AtRAC1 (ROP6) and ROP10, have been found to be negative regulators of stomatal ABA response. As a guard cell enriched ROP GTPase, AtRAC1 mediates stomatal ABA responses through regulation of the guard cell actin cytoskeleton. In transgenic plants expressing dominant-positive (i.e. constitutively active) AtRAC1 (DP-ROP), ABA-induced actin disruption as well as ABA-induced stomatal closure were impaired. These impaired phenotypes were also observed in the abi1-1 mutant, in which AtRAC1 cannot be activated by ABA. Furthermore, dominant negative AtRAC1 (DN-ROP) transgenic plants showed stomatal closure even without ABA treatment, both in the wild type background, and in an abi1-1 mutant background [176]. However, as discussed above, current evidence is not sufficient to infer the relationship between AtRAC1 and ABI1 in the wild type signaling network. Compared to AtRAC1, enrichment of ROP10 in guard cells is less pronounced, based on promoter GUS analysis. A null mutant of ROP10 is hypersensitive to ABA-induced stomatal closure and is more drought resistant than wild type. ROP10 protein localization at the guard cell membrane is weakly disrupted in an ABA-hypersensitive mutant, era1-2 [177].

Another small GTPase, ROP2, is related to the stomatal light response. Plants transformed with constitutively active ROP2 protein have large stomatal apertures under darkness and smaller stomatal apertures under light then wild type, while stomata of a dominant negative ROP2 mutant as well as a rop2 knockout mutant showed faster and larger opening under light.
Further, thermal imaging showed that constitutively activated ROP2 (CA-ROP2) plants have higher leaf temperature due to smaller stomatal apertures [178]. These results support a role of ROP2 in light responses of stomata, however, earlier results also suggest that ROP2 is involved in PA-induced ROS production in guard cell [102]. Because PA and ROS are positive regulators of ABA-induced stomatal closure, it can be postulated that ROP2 is also involved in stomatal ABA responses.

2.5.4 Farnesyl Transferase

ERA1 is the β-subunit of Arabidopsis farnesyl transferase [179]. In yeast and mammalian systems, farnesyl transferase catalyzes lipid modifications of signal transduction proteins to target them to the plasma membrane. Arabidopsis era1 mutants are hypersensitive to ABA-induced stomatal closure, partly due to reduced ABA activation of Ca\(^{2+}\) channels and anion channels in era1-2 guard cells [180, 181]. The era1-2 mutant has a reduced wilting phenotype under drought condition but also has reduced growth compared to wild type. The effect on crop drought tolerance of changing farnesyl transferase expression was tested in Brassica napus [182]. The drought-inducible promoter used in this study provides reversible and conditional expression of an antisense ERA1 construct, which overcomes the pleiotropic effects of constitutive silencing of ERA1. Seed yield was increased in the transgenic plants under drought conditions, but was not affected under normal growth condition [182]. The conditional knockout strategy can be applied to crops using other regulators of stomatal ABA response to improve crop drought tolerance.
2.5.5 Genes related to humidity sensing.

Genetic regulators of plant humidity responses have begun to emerge in recent years. Early experiments showed that two Arabidopsis MAP kinases, AtMPK4 and AtMPK6, were not responsive at the transcript level, but were activated by humidity and osmotic stresses [183]. Later, a protein kinase SnRK2e [184], subsequently called OST1 [39], was found to be not only activated by ABA but also by low humidity in plants. By reverse genetic approach, a T-DNA insertional mutant of SnRK2e was isolated and young mutant plants showed a wilting phenotype under rapid low humidity treatment. In the same study, it was found that several other protein kinases were activated by low humidity or ABA. The identities of these other proteins remain elusive [184]. OST1 was also identified, as was ABA2, from a forward genetic screen for humidity response mutants, using thermal imaging of leaf surface temperature on an EMS-mutagenized population [39]. Another ABA catabolism gene, CYP707A1, is also induced at the transcript level by high humidity in guard cells in Arabidopsis [54]. An mRNA cap binding protein mutant, abh1, also exhibits an altered response to humidity [185]. Under low humidity growth conditions, abh1 mutants have smaller stomatal apertures as compared to wildtype plants, while subsequent electrophysiology experiments correlated that phenotype to enhanced slow anion channel activity and reduced inward K$^+$ channel activity. The abh1 mutant is also hypersensitive to ABA-inhibited seed germination and has strong whole plant phenotypes [186]. The Arabidopsis anion channel mutant, slac1, also has altered humidity responses [27]. Another experiment showed that LCBK, a sphingoid long-chain base (LCB) kinase, is slightly activated by humidity [187]. As discussed above, sphingosine signaling has been found to be involved in stomatal ABA response [165]. All these experiments imply a connection between ABA signaling and humidity responses, however, none of these results can rule out the possibility of a humidity specific pathway that is independent from ABA signaling [188]. Finally, a novel
hat1 mutant was found to survive low humidity for 6 days while wildtype plants died in 24 hours. hat1 maps to a 168kb region in chromosome 5 which contains 21 genes [189]. Cloning of HAT1 may provide new insight into Arabidopsis humidity responses.

2.6. Transcriptional regulation in stomatal drought response.

In addition to inducing fast and reversible responses such as stomatal closure, drought stress also causes dynamic changes in gene expression patterns, which are considered to be slower than most of the above discussed signaling processes. Using microarray and sequencing technologies, thousands of gene expression changes were found in response to drought or ABA at the whole plant level [36, 190] and in guard cells [30]. A protein phosphatase 2C gene, AtPP2C-HA, was identified based on the results from the guard cell microarray [30]. In this section, several transcriptional regulators involved in stomatal drought response will be discussed. Because drought and other environmental factors can also modulate stomatal size and distributions in developing leaves, recent advances in the stomatal development and patterning will also be discussed in this section.

2.6.1 Transcription factors in stomatal drought response

ABF2 [191], ABF3 and ABF4 [192] are basic Leu zipper (bZip) proteins that can bind to ABA response elements (ABREs) of promoters. All three genes have strong expression in guard cells, suggesting functions in regulation of stomatal gene expression. Transgenic plants with overexpression of any of these three genes showed reduced water loss from excised leaves and increased drought tolerance, which is likely due to reduced stomatal opening and changes in expression of genes such as ABI1, ABI2, KAT1 and KAT2 [191, 192]. A RING finger E3 ligase,
SDIR1, is a putative upstream regulator of ABF3 and ABF4, because overexpression of ABF3 or ABF4 can partially rescue the ABA insensitive phenotype of an sdrl knockout mutant [193]. SDIR1 is expressed in all tissues tested and significantly induced by drought in guard cells. Plants with SDIR1 overexpression are more tolerant to drought while sdrl knockout plants are more sensitive to drought as compared to wild type. Under normal watering, SDIR1 overexpression lines also have smaller stomatal apertures, and are hypersensitive to ABA-induced stomatal closure, while knockout lines have larger stomata and are less sensitive to ABA compared to wild type and sdrl knockout lines [193].

Three R2R3-Myb transcription factors, AtMYB44, AtMYB60 and AtMYB61, are also related to stomatal function. All three Myb transcription factors are highly expressed in guard cells as shown by promoter reporter gene assays. However, each of these three genes acts on different aspects of stomatal responses. AtMYB44 overexpression lines have smaller stomatal apertures in normal conditions and faster closure rate under ABA treatment as compared to wild type, while knocking out AtMYB44 does not result in altered phenotypes [194]. Unlike AtMYB44, AtMYB60 and AtMYB61 are not involved in stomatal ABA responses, but in light regulation of stomatal movements. T-DNA insertional mutant atmyb60 has smaller stomatal opening under light as compared to wild type, while a myb61 mutant showed defects in dark-induced stomatal closure [195, 196]. Drought stress responses were directly tested for AtMYB60 and AtMYB44. Both the atmyb60 mutant and AtMYB44 over-expression plants showed increased drought tolerance, which is consistent with their roles in stomatal regulation [194, 195].

ABO1 is homologous to a subunit in yeast and human elongator complex, which is a multi-protein complex functioning in transcript elongation. abol-1 mutants lose water more slowly than wild-type plants, and are hypersensitive to ABA-induced stomatal closure. The abol-1 mutation also affects stomatal development, because almost half of identified guard cell pairs form immature stomata with closed apertures [197].
2.6.2. Stomatal development and drought responses.

Stomata develop from leaf protodermal cells through multiple steps including stomatal spacing divisions, guard mother cell division and guard cell maturation. Stomatal spacing determines stomatal distribution and density, while stomatal sizes are determined by guard cell maturation [198]. Sizes and distribution of stomatal complexes on mature leaf surfaces are fixed, restricting acclimation to environmental changes to the changes of stomatal apertures. However, both the sizes and distribution of stomata as determined during leaf maturation limit the capacity of stomatal conductance [199]. Many environmental signals can modify the density and sizes of stomata during leaf development. Under water limiting conditions, cotton plants develop smaller and denser stomata in contrast to well watered conditions [200], and treatment of *Tradescantia virginiana* plants with ABA causes similar phenotype [201]. Both the effects of CO₂ and humidity on stomatal development are blocked in an ABA biosynthesis mutant, *aba1*, implying that both regulatory effects are mediated by ABA [202].

A number of genetic regulators of stomatal patterning have been discovered in recent years. A subtilisin like serine protease, SDD1 and a secretory peptide [203, 204], EPIDERMAL PATTERNING FACTOR 1 (EPF1) [205], were hypothesized to be involved in extracellular signaling during stomatal development. Transmembrane receptor-like protein TMM [206], and transmembrane receptor-like kinases, ER, ERL1, ERL2 [207], are intermediate components between extracellular signals and intracellular signaling pathways of stomatal development. A MAPK cascade, including YODA, MKK4/MKK5 and MPK3/MPK6, acts downstream of membrane receptor kinases to regulate stomatal development [198, 208]. One member in this cascade, MPK3, is also a positive regulator of stomatal ABA and ROS responses [134]. Downstream of this MAPK cascade, a subfamily of basic helix-loop-helix (bHLH) transcription factors, SPCH, MUTE and FAMA, are involved in the control of stomatal lineage transitions.
Another subfamily of bHLH transcription factors, ICE1 and SCRM2 form dimers with SPCH, MUTE and FAMA, and thus control stomatal initiation, proliferation and differentiation [213]. Interestingly, ICE1 was previously found to be important for plant cold tolerance [214]. Finally, two genes, FLPS and Myb88, which are paralogous proteins of one subfamily of Myb transcription factors, also affect stomatal development [215]. Because both ICE1 and MPK3 are also involved in plant response to environmental signals, these two genes are potential candidates for further analysis of how environmental factors modulate stomatal developmental programs [216].

2.7. Summary

Stomata are microscopic pores on leaf surfaces flanked by pairs of guard cells. Guard cells responded to a range of different environmental stimuli, such as light, humidity, temperature and CO2 concentration, and also a number of internal stimuli mediated by plant hormones. The main role of stomata in plant drought response is to limit transpirational water loss, through reducing stomatal apertures in response to the stress hormone ABA. In *Arabidopsis*, many positive and negative regulators of guard cell drought responses have been identified by genetic and transgenic approaches. Mutants in these mediators sometimes showed improved plant drought tolerance, mainly due to increased guard cell ABA sensitivity (see Table 1.1) or increased ABA biosynthesis. However, as a general stress hormone, ABA also has other functions such as inhibiting seed germination and restricting plant growth. These traits, especially ABA inhibition of seed germination, are observed in many mutants that have hypersensitive stomatal ABA responses. ABA hypersensitive traits in organs other than guard cells could be detrimental to crop plants. Mutants with defects in light induced stomatal opening could also have better drought tolerance due to their intrinsically smaller stomatal apertures under light. However,
the photosynthesis rate and total assimilation, although not measured in most published papers, is likely to be reduced in these mutants, because the reduced apertures can limit CO₂ diffusion into leaves. Therefore, more analyses of transpiration efficiency are needed to assess the utility of these mutations for crop improvement.

Some genes, for example ER-family genes, simultaneously modify stomatal development and other aspects of leaf morphology, which both modulate plant water use efficiency. Genetic regulators with similar dual or multiple functions could be potential targets for further experimental analysis and for crop breeding. Arabidopsis heterotrimeric G proteins are good candidates, because G protein mutants, *gpa1* and *agb1*, have both altered stomatal ABA responses, stomatal density changes and leaf morphology changes. G protein mutants are hypersensitive to ABA inhibition of germination and root growth, but hyposensitive to ABA-inhibited stomatal opening [217]. Such opposite functions of G proteins imply that wild type plants have longer root growth, better germination rate and better stomatal response as compared to mutants that lack G protein components. Thus, it can be postulated that G proteins are global positive regulators of plant growth.

Despite the enormous progress made in understanding the genetic regulators of stomatal drought response in *Arabidopsis*, we are far from understanding the complex interactions between these regulators. To apply our knowledge in *Arabidopsis* in improving crop drought tolerance, one strategy would be generating over-expression or RNAi transgenic plant for all the known regulators and then testing these plants under identical conditions, which is unfeasible for most other species. The alternative strategy would be to first synthesize our knowledge about the interconnection of these regulators into a network, and then use computational simulation to test the synthesized network until it can reproduce the observed interactions and responses in stomata [5]. We also need to keep in mind that those key regulators discovered in *Arabidopsis*, may have altered their function or gained functional redundancy in other species. As more fully sequenced
plant genomes become available, comparative genomics can also help us to identify conserved
genes that are universally involved in stomatal functions.

Because drought is a complex stress interweaving with many other stresses, studies of
crosstalk are also of great importance for practical concerns such as crop breeding. During the last
two decades, the Arabidopsis community has accumulated an unprecedented number of well
characterized gene-centric mutants involved in both drought response and other aspects of plant
physiology. When studying stomatal drought responses, these well characterized mutants are
valuable biological tools for understanding crosstalk between cellular processes under drought or
between drought and other stresses.
Chapter 3 Predicting essential components of signal transduction networks: a dynamic model of guard cell abscisic acid signaling

This chapter is previously published as a research article [5], and was written specifically to fulfill thesis requirements.

3.1 Abstract

Plants both lose water and take in carbon dioxide through microscopic stomatal pores, each of which is regulated by a surrounding pair of guard cells. During drought, the plant hormone abscisic acid (ABA) inhibits stomatal opening and promotes stomatal closure, thereby promoting water conservation. Dozens of cellular components have been identified to function in ABA regulation of guard cell volume and thus of stomatal aperture, but a dynamic description is still not available for this complex process. Here we synthesize experimental results into a consistent guard cell signal transduction network for ABA-induced stomatal closure, and develop a dynamic model of this process. Our model captures the regulation of more than 40 identified network components, and accords well with previous experimental results at both the pathway and whole-cell physiological level. By simulating gene disruptions and pharmacological interventions we find that the network is robust against a significant fraction of possible perturbations. Our analysis reveals the novel predictions that the disruption of membrane depolarizability, anion efflux, actin cytoskeleton reorganization, cytosolic pH increase, the phosphatidic acid pathway, or K⁺ efflux through slowly activating K⁺ channels at the plasma membrane lead to the strongest reduction in ABA responsiveness. Initial experimental analysis assessing ABA-induced stomatal closure in the presence of cytosolic pH clamp imposed by the
weak acid butyrate is consistent with model prediction. Simulations of stomatal response as derived from our model provide an efficient tool for the identification of candidate manipulations that have the best chance of conferring increased drought stress tolerance and for the prioritization of future wet bench analyses. Our method can be readily applied to other biological signaling networks to identify key regulatory components in systems where quantitative information is limited.

3.2 Introduction

One central challenge of systems biology is the distillation of systems level information into applications such as drug discovery in biomedicine or genetic modification of crops. In terms of applications it is important and practical that we identify the subset of key components and regulatory interactions whose perturbation or tuning leads to significant functional changes (e.g., changes in a crop’s fitness under environmental stress or changes in the state of malfunctioning cells, thereby combating disease). Mathematical modeling can assist in this process by integrating the behavior of multiple components into a comprehensive model that goes beyond human intuition, and also by addressing questions that are not yet accessible to experimental analysis.

In recent years, theoretical and computational analysis of biochemical networks has been successfully applied to well defined metabolic pathways, signal transduction, and gene regulatory networks [218-220]. In parallel, high-throughput experimental methods have enabled the construction of genome-scale maps of transcription factor-DNA and protein-protein interactions [221, 222]. The former are quantitative, dynamic descriptions of experimentally well-studied cellular pathways with relatively few components, while the latter are static maps of potential interactions with no information about their timing or kinetics. Here we introduce a novel approach that stands in the middle ground of the abovementioned methods by incorporating
the synthesis and dynamic modeling of complex cellular networks that contain diverse, yet only qualitatively known regulatory interactions.

We develop a mathematical model of a highly complex cellular signaling network and explore the extent to which the network topology determines the dynamic behavior of the system. We choose to examine signal transduction in plant guard cells for two reasons. First, guard cells are central components in control of plant water balance, and better understanding of their regulation is important for the goal of engineering crops with improved drought tolerance. Second, abscisic acid (ABA) signal transduction in guard cells is one of the best characterized signaling systems in plants: more than 20 components, including signal transduction proteins, secondary metabolites, and ion channels, have been shown to participate in ABA-induced stomatal closure. ABA induces guard cell shrinkage and stomatal closure via two major secondary messengers, cytosolic Ca\(^{2+}\) (Ca\(^{2+}\)) and cytosolic pH (pH\(_c\)). A number of signaling proteins and secondary messengers have been identified as regulators of Ca\(^{2+}\) influx from outside the cell or Ca\(^{2+}\) release from internal stores; the downstream components responding to Ca\(^{2+}\) are certain vacuolar and plasma membrane K\(^+\) permeable channels, and anion channels in the plasma membrane [95, 223]. Increases in cytosolic pH promote the opening of anion efflux channels and enhance the opening of voltage-activated outward K\(^+\) channels in the plasma membrane [25, 224, 225]. Stomatal closure is caused by osmotically driven cell volume changes induced by both K\(^+\) and anion efflux through plasma membrane-localized channels. Despite the wealth of information that has been collected regarding ABA signal transduction, the majority of the regulatory relationships is known only qualitatively and is studied in relative isolation, without considering their possible feedback or crosstalk with other pathways. Therefore, in order to synthesize this rich knowledge, one needs to assemble the information on regulatory mechanisms involved in ABA-induced stomatal closure into a system-level regulatory network that is consistent with experimental observations. Clearly, it is difficult to assemble the network and predict the dynamics of this
system from human intuition alone, and thus theoretical tools are needed.

We synthesize the experimental information available about the components and processes involved in ABA-induced stomatal closure into a comprehensive network, and study the topology of paths between signal and response. To capture the dynamics of information flow in this network we express synergy between pathways as combinatorial rules for the regulation of each node, and formulate a dynamic model of ABA-induced closure. Both in silico and in initial experimental analysis, we study the resilience of the signaling network to disruptions. We systematically sample functional and dynamic perturbations in network components and uncover a rich dynamic repertoire ranging from ABA hypersensitivity to complete insensitivity. Our model is validated by its agreement with prior experimental results, and yields a variety of novel predictions that provide targets on which further experimental analysis should focus. To our knowledge, this is one of the most complex biological networks ever modeled in a dynamical fashion.
3.3 Results

3.3.1 Extraction and Organization of Data from the Literature

We focus on ABA induction of stomatal closure, rather than ABA inhibition of stomatal opening, because these two processes, although related, exhibit distinct mechanisms, and there is substantially more information on the former process than on the latter in the literature. Experimental information about the involvement of a specific component in ABA-induced stomatal closure can be partitioned into three categories. First, biochemical evidence provides information on enzymatic activity or protein-protein interactions. For example, the putative G protein-coupled receptor 1 (GCR1) can physically interact with the heterotrimeric G protein a component 1 (GPA1) as supported by split-ubiquitin and coimmunoprecipitation experiments [167]. Second, genetic evidence of differential responses to a stimulus in wild-type plants versus mutant plants implicates the product of the mutated gene in the signal transduction process. For example, the ethyl methanesulfonate-generated ost1 mutant is less sensitive to ABA; thus, one can infer that the OST1 protein is a part of the ABA signaling cascade [226]. Third, pharmacological experiments, in which a chemical is used either to mimic the elimination of a particular component, or to exogenously provide a certain component, can lead to similar inferences. For example, a nitric oxide (NO) scavenger inhibits ABA-induced closure, while a NO donor promotes stomatal closure; thus, NO is a part of the ABA network [61]. The last two types of inference do not give direct interactions but correspond to pathways and pathway regulation. The existing theoretical literature on signaling is focused on networks where the first category of information is known, along with the kinetics of each interaction. However, the availability of such detailed knowledge is very much the exception rather than the norm in the
experimental literature. Here we propose a novel method of representing qualitative and incomplete experimental information and integrating it into a consistent signal transduction network.

First, we distill experimental conclusions into qualitative regulatory relationships between cellular components (signaling proteins, metabolites, ion channels) and processes. For example, the evidence regarding OST1 and NO is summarized as both OST1 and NO promoting ABA-induced stomatal closure. We distinguish between positive and negative regulation by using the verbs “promote” and “inhibit” represented graphically as “→” and “—l” respectively, and quantify the severity of the effect by the qualifier “partial” A partial promoter’s (inhibitor’s) loss has less severe effects than the loss of a promoter (inhibitor), most probably due to other regulatory effects on the target node. Using these relations, we construct a database that contains more than 140 entries and is derived from more than 50 literature citations on ABA regulation of stomatal closure (Appendix A. Table S1). A number of entries in the database correspond to a component-to-component relationship, such as “A promotes B” which is mostly obtained by pharmacological experiments (e.g., applying A causes B response). However, the majority of the entries belong to the two categories of indirect inference described above, and are of the type “C promotes the process (A promotes B).” This kind of information can be obtained from both genetic and pharmacological experiments (e.g., disrupting C causes less A-induced B response, or applying C and A simultaneously causes a stronger B response than applying A only). There are a few instances of documented independence of two cellular components, which we identify with the qualifier “no relationship.” Most of the information is derived from the model species Arabidopsis thaliana, but data from other species, mostly Vicia faba, are also included where comparable information from Arabidopsis thaliana is lacking.
3.3.2 Assembly of the ABA Signal Transduction Network

To synthesize all this information into a consistent network, we need to determine how the different pathways suggested by experiments fit together (i.e., we need to find the pathways’ branching and crossing points). We develop a set of rules compatible with intuitive inference, aiming to determine the sparsest graph consistent with all experimental observations. We summarize the most important rules in Figure 3.1; in the following we give examples for their application.

![Figure 3.1](https://example.com/figure3.1.png)

Figure 3.1. Illustration of the Inference Rules Used in Network Reconstruction

1. If \( A \rightarrow B \) and \( C \rightarrow \) process (\( A \rightarrow B \)), where \( A \rightarrow B \) is not a biochemical reaction such as an enzyme catalyzed reaction or protein-protein/small molecule interaction, we assume that \( C \) is acting on an intermediary node (IN) of the A–B pathway.

2. If \( A \rightarrow B, A \rightarrow C, \) and \( C \rightarrow \) process (\( A \rightarrow B \)), where \( A \rightarrow B \) is not a direct interaction, the most parsimonious explanation is that \( C \) is a member of the A–B pathway, i.e. \( A \rightarrow C \rightarrow B \).

3. If \( A \rightarrow! B \) and \( C \rightarrow! \) process (\( A \rightarrow! B \)), where \( A \rightarrow! B \) is not a direct interaction, we assume that \( C \) is inhibiting an intermediary node (IN) of the A–B pathway. Note that \( A \rightarrow! IN \rightarrow! B \) is the only logically consistent representation of the A–B pathway.

If \( A \rightarrow B \) and \( C \rightarrow \) process (\( A \rightarrow B \)), where \( A \rightarrow B \) is not a biochemical reaction such as an enzyme catalyzed reaction or protein–protein/small molecule interaction, we assume that \( C \) is acting on an intermediary node (IN) of the A–B pathway. This IN could be an intermediate protein complex, protein-small molecule complex, or multiple complexes (see Figure 3.1, panel 1). For example, ABA → closure, and NO synthase (NOS) → process (ABA → closure); therefore, ABA → IN → closure, NOS → IN. If \( A \rightarrow B \) is a direct process such as a biochemical
reaction or a protein-protein interaction, we assume that C → process (A → B) corresponds to C → A → B.

A → B and C → process (A → B) can be transformed to A → C → B if A → C is also documented. This means that the simplest explanation is to identify the putative intermediary node with C. For example, ABA → NOS, and NOS → process (ABA → NO) are experimentally verified and NOS is an enzyme producing NO, therefore, we infer ABA → NOS → NO (see Figure 3.1, panel 2).

A rule similar to rule 1 applies to inhibitory interactions (denoted by —↓); however, in the case of A —↓ B, and C —↓ process (A —↓ B), the logically correct representation is: A → IN —↓ B, C —↓ IN (see Figure 3.1, panel 3).

The above rules constitute a heuristic algorithm for first expanding the network wherever the experimental relationships are known to be indirect, and second, minimizing the uncertainty of the network by filtering synonymous relationships. Mathematically, this algorithm is related to the problem of finding the minimum transitive reduction of a graph (i.e., for finding the sparsest subgraph with the same reachability relationships as the original) [227]; however, it differs from previously used algorithms by the fact that the edges can have one of two signs (activating and inhibitory), and edges corresponding to direct interactions are maintained.

In the reconstructed network, given in Figure 3.2, the network input is ABA and the output is the node “Closure.” The small black filled circles represent putative intermediary nodes mediating indirect regulatory interactions. The edges (lines) of the network represent interactions and processes between two components (nodes); an arrowhead at the end of an edge represents activation, and a short segment at the end of an edge signifies inhibition. Edges that signify interactions derived from species other than Arabidopsis are colored light blue. We indicate two inferred negative feedback loops on S1P and pH, (see below) by dashed light blue lines. Nodes involved in the same metabolic reaction or protein complex are bordered by a gray box; only
those arrows that point into or out of the box signify information flow (signal transduction). Some of the edges on Figure 3.2 are not explicitly incorporated in Appendix A. Table S1 because they represent general biochemical or physical knowledge (e.g., reactions inside gray boxes or depolarization caused by anion efflux).

Figure 3.2. Current Knowledge of Guard Cell ABA Signaling
The color of the nodes represents their function: enzymes are shown in red, signal transduction proteins are green, membrane transport–related nodes are blue, and secondary messengers and small molecules are orange. Small black filled circles represent putative intermediary nodes mediating indirect regulatory interactions. Arrowheads represent activation, and short perpendicular bars indicate inhibition. Light blue lines denote interactions derived from species other than Arabidopsis; dashed light-blue lines denote inferred negative feedback loops on pHc and S1P. Nodes involved in the same metabolic pathway or protein complex are bordered by a gray box; only those arrows that point into or out of the box signify information flow (signal transduction).

The full names of network components corresponding to each abbreviated node label are: ABA, abscisic acid; ABI1/2, protein phosphatase 2C ABI1/2; ABH1, mRNA cap binding protein; Actin, actin cytoskeleton reorganization; ADPRc, ADP ribose cyclase; AGB1, heterotrimeric G protein
β component; AnionEM, anion efflux at the plasma membrane; Arg, arginine; AtPP2C, protein phosphatase 2C; AtRboh, NADPH oxidase; CaM, Ca$^{2+}$ influx across the plasma membrane; Ca$^{2+}$ ATPase, Ca$^{2+}$ ATPases and Ca$^{2+}$/H$^{+}$ antiporters responsible for Ca$^{2+}$ efflux from the cytosol; Ca$^{2+}_{c}$, cytosolic Ca$^{2+}$ increase; cADPR, cyclic ADP-ribose; cGMP, cyclic GMP; CIS, Ca$^{2+}$ influx to the cytosol from intracellular stores; DAG, diacylglycerol; Depolar, plasma membrane depolarization; ERA1, farnesyl transferase ERA1; GC, guanly cyclase; GCR1, putative G protein–coupeld receptor; GPA1, heterotrimeric G protein α subunit; GTP, guanosine 5’-triphosphate; H$^{+}$ ATPase, H$^{+}$ ATPase at the plasma membrane; InsPK, inositol polyphosphate kinase; InsP3, inositol-1,4,5-trisphosphate; InsP6, inositol hexakisphosphate; KAP, K$^{+}$ efflux through rapidly activating K$^{+}$ channels (AP channels) at the plasma membrane; KEV, K$^{+}$ efflux from the vacuole to the cytosol; KOUT, K$^{+}$ efflux through slowly activating outwardly-rectifying K$^{+}$ channels at the plasma membrane; NAD$, nicotinamide adenine dinucleotide; NADPH, nicotinamide adenine dinucleotide phosphate; NOS, Nitric oxide synthase; NIA12, Nitrate reductase; NO, Nitric oxide; OST1, protein kinase open stomata 1; PA, phosphatidic acid; PC, phosphatidyl choline; PEPC, phosphoenolpyruvate carboxylase; PIP2, phosphatidylinositol 4,5-bisphosphate; PLC, phospholipase C; PLD, phospholipase D; RAC1, small GTPase RAC1; RCN1, protein phosphatase 2A; ROP2, small GTPase ROP2; ROP10, small GTPase ROP10; ROS, reactive oxygen species; SphK, sphingosine kinase; S1P, sphingosine-1-phosphate.

A brief biological description of this reconstructed network (Figure 3.2) is as follows. ABA induces guard cell shrinkage and stomatal closure via two major secondary messengers, Ca$^{2+}_{c}$ and pH$_{c}$. Two mechanisms of Ca$^{2+}_{c}$ increase have been identified: Ca$^{2+}$ influx from outside the cell and Ca$^{2+}$ release from internal stores. Ca$^{2+}$ can be released from stores by InsP3 [228] and InsP6 [229], both of which are synthesized in response to ABA, or by cADPR and cGMP [107], whose upstream signaling molecule, NO [61, 230], is indirectly activated by ABA. Opening of channels mediating Ca$^{2+}$ influx is mainly stimulated by reactive oxygen species (ROS) [99], and we reconstruct two ABA-ROS pathways involving OST1 [226] and GPA1 (L. Perfus-Barbeoch and S. M. Assmann, unpublished data), respectively. Based on current experimental evidence these two pathways are distinct, but not independent. The downstream components responding to Ca$^{2+}$ are certain vacuolar and plasma membrane K$^{+}$ permeable channels, and anion channels in the plasma membrane [95, 223]. The mechanism of pH control by ABA is less clear, but it is known that pH$_{c}$ increases shortly after ABA treatment [72, 231]. Increases in pH$_{c}$ levels promote the opening of anion efflux channels and enhance the opening of voltage-activated outward K$^{+}$ channels in the plasma membrane [25, 224, 225]. Stomatal closure is caused by osmotically
driven cell volume changes induced by K\(^+\) and anion efflux through plasma membrane-localized channels, and there is a complex interregulation between ion flux and membrane depolarization.

In addition to the secondary-messenger-induced pathways, there are two less-well-studied ABA signaling pathways involving the reorganization of the actin cytoskeleton, and the organic anion malate. ABA inactivates the small GTPase protein RAC1, which in turn blocks actin cytoskeleton disruption [176], contributing to an ABA-induced actin cytoskeleton reorganization process that is potentially Ca\(^{2+}\) dependent [232]. In our model system, *Arabidopsis*, ABA regulation of malate levels has not been described. However, in *V. faba* it has been shown that ABA inhibits PEP carboxylase and malate synthesis [233], and that ABA induces malate breakdown [234]. In some conditions sucrose is an osmoticum that contributes to guard cell turgor [235, 236] but no mechanisms of ABA regulation of sucrose levels have been described.

The recessive mutant of the protein phosphatase 2C (PP2C) ABI1, *abi1-1R*, is hypersensitive to ABA [144, 145]. ABI1 is negatively regulated by phosphatidic acid (PA) and ROS, and pH, can activate ABI1 [104, 146, 148]. ABI1 negatively regulates RAC1 [176]. We hypothesize that ABI1 negatively regulates the NADPH oxidase (Atrboh) because ABI1 negatively regulates ROS production and Atrboh has been shown to be the dominant producer of ROS in guard cells [237]. We also assume that ABI1 inhibits anion efflux at the plasma membrane, because the dominant *abi1-1* mutant is known to affect the ABA response of anion channels [238] and because anion channels are documented key regulators of ABA-induced stomatal closure [239]. Components functioning downstream from ABI2 and its role in guard cell signaling are not well established, so ABI2 is not included. The newly isolated PP2C recessive mutants, *AtP2C-HA* [30] and *AtPP2CA* [153], exhibit minor ABA hypersensitivity. However, their downstream targets remain elusive; thus, we incorporate them as a general inhibitor of closure denoted AtPP2C.

Mutation of the gene encoding the mRNA cap-binding protein, ABH1, results in
hypersensitivity of ABA-induced $\text{Ca}^{2+}$ elevation/oscillation and of anion efflux in plants grown under some environmental conditions [185, 186]. We assume an inhibitory effect of ABH1 on $\text{Ca}^{2+}$ influx across the plasma membrane (CaM), which can explain both of these effects due to the $\text{Ca}^{2+}$ regulation of anion efflux. Since the abh1 mutation affects transcript levels of some genes involved in ABA response, this mutation may also affect ABA sensitivity by altering gene expression rather than by regulation of the rapid signaling events on which our network focuses. Mutations in the gene encoding the farnesyl transferase ERA1 or the gene encoding GCR1 also lead to hypersensitive ABA-induced closure; ERA1 has been shown to negatively regulate CaM and anion efflux [180, 181], whereas GCR1 has been shown to be interact with GPA1 [167]. We assume that ERA1 negatively regulates CaM and GCR1 negatively regulates GPA1.

Another assumption in the network is that the protein phosphatase RCN1/PP2A regulates nitrate reductase (NIA12) activity as observed in spinach leaf tissue; this is expected to be a well-conserved mechanism due to the high sequence conservation of NIA-PP2A regulatory domains [240]. Figure 3.2 contains two putative autoregulatory negative feedback loops acting on S1P and $\text{pH}_c$, respectively. The existence of feedback regulation can be inferred from the published timecourse measurements of S1P [165] and $\text{pH}_c$ [72] — both indicating a fast increase in response to ABA, then a decrease — but the mediators are currently unknown. The assembled network is consistent with our biological knowledge with minimal additional assumptions, and it will serve as the starting point for the graph analysis and dynamic modeling described in the following sections.

3.3.3 Modeling ABA Signal Transduction

Signaling networks can be represented as directed graphs where the orientation of the edges reflects the direction of information propagation (signal transduction). In a signal
transduction network there exists a clear starting point, the node representing the signal (here, ABA), and one can follow the paths (successions of edges) from that starting point to the node(s) representing the output(s) of the network (here, stomatal closure). The signal-output paths correspond to the propagation of reactions in chemical space, and can be thought of as pseudodynamics [241]. When only static information is available, pseudodynamics takes into account the graph theoretical properties of the signal transduction network. For example, one can measure the number of nodes or distinct network motifs that appear one, two,…n edges away from the signal node. Such motifs reflect different cellular signaling processing capabilities and provide important insights into the biological processes under investigation. Graph theoretical measures can also provide information about the importance (centrality) of signal mediators [8] and can predict the changes in path structure when nodes or edges in the network are disrupted. These disruptions, explored experimentally by genetic mutations, voltageclamping, or pharmacological interventions, can be modeled in silico by removing the perturbed node and all its edges from the graph [8]. The absence of nodes and edges will disrupt the paths in the network, causing a possible increase in the length of the shortest path between signal (ABA) and output (closure), suggesting decreased ABA sensitivity, or in severe cases the loss of all paths connecting input and output (i.e., ABA insensitivity).

We find that there are several partially or completely independent (nonoverlapping) paths between ABA and closure. The path of pH-induced anion efflux is independent of the paths involving changes in Ca\(^{2+}\). Based on the current knowledge incorporated in Figure 3.2, the path mediated by malate breakdown is independent of both Ca\(^{2+}\) and pH signaling. This result could change if evidence of a suggested link between pH and malate regulation [242] is found; note that regulation of malate synthesis in guard cells appears to have cell-specific aspects [243]. Increase in Ca\(^{2+}\) can be induced by several independent paths involving ROS, NO, or InsP6. Thanks to the existence of numerous redundant signal (ABA)-output (closure) paths, a complete disconnection
of signal from output (loss of all the paths) is possible only if four nodes, corresponding to actin reorganization, pH Increase, malate breakdown, and membrane depolarization, are simultaneously disrupted. This indicates a remarkable topological resilience, and suggests that functionally redundant mechanisms can compensate for single gene disruptions and can maintain at least partial ABA sensitivity. However, path analysis alone cannot capture bidirectional signal propagation and synergy (cooperativity) in living biological systems. For example, two nonoverlapping paths that reach the node closure could be functionally synergistic. Using only path analysis, disruption of either path would not be predicted to lead to a disconnection of the signal (ABA) from the output (closure), but due to the synergy between the two paths, the closure response may be strongly impaired if either of the two paths is disrupted experimentally. Because of such limitations of path analysis, we turn from path analysis to a dynamic description.

Dynamic models have as input information (1) the interactions and regulatory relationships between components (i.e., the interaction network); (2) how the strength of the interactions depends on the state of the interacting components (i.e., the transfer functions); and (3) the initial state of each component in the system. Given these, the model will output the time evolution of the state of the system (e.g., the system’s response to the presence or absence of a given signal). Given the incomplete characterization of the processes involved in ABA-induced stomatal closure (as is typical of the current state of knowledge of cell signaling cascades), we employ a qualitative modeling approach. We assume that the state of the network nodes can have two qualitative values: 0 (inactive/off) and 1 (active/on) [244]. These values can also describe two conformational states of a protein, such as closed and open states of an ion channel, or basal and high activity for enzymes. This assumption is necessary due to the absence of quantitative concentration or activity information for the vast majority of the network components. It is additionally justified by the fact that in the case of combinatorial regulation or cooperative binding, the input-output relationships are sigmoidal and thus can be distilled into two discrete
output states [245].

Since “stomatal closure” does not usually entail the complete closure of the stomatal pore but rather a clear decrease in the stomatal aperture, and since there is a significant variability in the response of individual stomata, the threshold separating the off (0) and on (1) state of the node “Closure” needs to invoke a population level description. We measured the stomatal aperture size distribution in the absence of ABA or after treatment with 50 μM ABA (see Materials and Methods). Our first observation was the population-level heterogeneity of stomatal apertures even in their resting condition (Figure 3.3A), a fact that may not be widely appreciated when more standard presentations, such as mean ± standard error, are used (see Figure 3.3B). The stomatal aperture distribution shifts towards smaller apertures after ABA treatment, and also broadens considerably. The latter result is inconsistent with the assumption of each stomate changing its aperture according to a common function that decreases with increasing ABA concentration, and suggests considerable cell-to-cell variation in the degree of response to ABA. Moreover, although there is a clear difference between the most probable “open” (0 ABA) and “closed” (+ ABA) aperture sizes, there also exists an overlap between the aperture size distribution of “open” and “closed” stomata. This result indicates the possibility of differential and cell-autonomous stomatal responses to ABA. In the absence of 6 ABA measurements on the same stomate, we define the threshold of closure as a statistically significant shift of the stomatal aperture distribution towards smaller apertures in response to ABA signal transduction.
Figure 3.3. Stomatal Aperture Distributions without ABA Treatment (graybars) and with 50 µM ABA (white bars)
(A) The x axis gives the stomatal aperture size and the y axis indicates the fraction of stomata for which that aperture size was observed. The black columns indicate the overlap between the 0 µM ABA and the 50 µM ABA distributions.
(B) Classical bar plot representation of stomatal aperture for treatment with 50 µM ABA (white bar, labeled 1) and without ABA treatment (gray bar, labeled 2) using mean ± standard error. This representation provides minimal information on population structure.

In our model the dynamics of state changes are governed by logical (Boolean) rules giving the state transition of each node given the state of its regulators (upstream nodes). We determine the Boolean transfer function for each node based on experimental evidence. The state of a node regulated by a single upstream component will follow the state of its regulator with a delay. If two or more pathways can independently lead to a node’s activation, we combine them with a logical “or” function. If two pathways cannot work independently, we model their synergy as a logical “and” function. For nodes regulated by inhibitors we assume that the necessary condition of their activation (state 1) is that the inhibitor is inactive (state 0). As all putative intermediary nodes of Figure 3.2 are regulated by a single activator, and regulate a single downstream component, they only affect the time delays between known nodes; for this reason we do not explicitly incorporate intermediary nodes as components of the dynamic model. Table 3.1 lists the regulatory rules of known nodes of Figure 3.2; we give a detailed justification of each rule in Appendix A. Text S1.
Table 3.1. Boolean Rules Governing the States of the Known (Named) Nodes in the Signal Transduction Network

<table>
<thead>
<tr>
<th>Node</th>
<th>Boolean Regulatory Rule</th>
</tr>
</thead>
<tbody>
<tr>
<td>NO</td>
<td>NO* = NIA12 and NOS</td>
</tr>
<tr>
<td>PLC</td>
<td>PLC* = ABA and Ca^{2+}_c</td>
</tr>
<tr>
<td>CaIM</td>
<td>CaIM* = (ROS or not ERA1 or not ABH1) and not Depolar</td>
</tr>
<tr>
<td>GPA1</td>
<td>GPA1* = (S1P or not GCR1) and AGB1</td>
</tr>
<tr>
<td>Atrboh</td>
<td>Atrboh* = pH_c and OST1 and ROP2 and not ABI1</td>
</tr>
<tr>
<td>H’ ATPase</td>
<td>Ht ATPase* = not ROS and not pH_c and not Ca^{2+}_c</td>
</tr>
<tr>
<td>Malate</td>
<td>Malate* = PEPC and not ABA and not AnionEM</td>
</tr>
<tr>
<td>RAC1</td>
<td>RAC1* = not ABA and not ABI1</td>
</tr>
<tr>
<td>Actin</td>
<td>Actin* = Ca^{2+}_c or not RAC1</td>
</tr>
<tr>
<td>ROS</td>
<td>ROS* = ABA and PA and pH_c</td>
</tr>
<tr>
<td>ABI1</td>
<td>ABI1* = pH_c and not PA and not ROS</td>
</tr>
<tr>
<td>KAP</td>
<td>KAP* = (not pH_c or not Ca^{2+}_c) and Depolar</td>
</tr>
<tr>
<td>Ca^{2+}_c</td>
<td>Ca^{2+}_c* = (CaIM or CIS) and not Ca^{2+}_c ATPase</td>
</tr>
<tr>
<td>CIS</td>
<td>CIS* = (cGMP and cADPR) or (InsP3 and InsP6)</td>
</tr>
<tr>
<td>AnionEM</td>
<td>AnionEM* = ((Ca^{2+}_c or pH_c) and not ABI1 ) or (Ca^{2+}_c or pH_c) and not ABI1 )</td>
</tr>
<tr>
<td>KOUT</td>
<td>KOUT* = (pH_c or not ROS or not NO) and Depolar</td>
</tr>
<tr>
<td>Depolar</td>
<td>Depolar* = KEV or AnionEM or not Ht ATPase or not KOUT or Ca^{2+}_c</td>
</tr>
<tr>
<td>Closure</td>
<td>Closure* = (KOUT or KAP ) and AnionEM and Actin and not Malate</td>
</tr>
</tbody>
</table>

The nomenclature of the nodes is given in the caption of Figure 3.2. The nodes that have only one input are not listed to save space; a full description and justification can be found in Appendix A. Text S1. The next state of the node on the left-hand side of the equation (marked by *) is determined by the states of its effector nodes according to the function on the right-hand side of the equation.

Frequently in Boolean models time is quantized into regular intervals (timesteps), assuming that the duration of all activation and decay processes is comparable [246]. For generality we do not make this assumption, and in the absence of timing or duration information we follow an asynchronous method that allows for significant stochasticity in process durations [247, 248]. Choosing as a timestep the longest duration required for a node to respond to a change in the state of its regulator(s) (also called a round of update, as each component’s state will be updated during this time interval), the Boolean updating rules of an asynchronous algorithm can be written as:

\[ S_i^n = B_i \left( S_j^{n_j}, S_k^{n_k}, S_l^{n_l}, ... \right), \]  

(1)
where $S_i^n$ is the state of component $i$ at timestep $n$, $B_i$ is the Boolean function associated with the node $i$ and its regulators $j, k, l, \ldots$ and $m_j, m_k, m_l, \ldots \in \{n-1, n\}$, signifying that the timepoints corresponding to the last change in a input node’s state can be in either the previous or current round of updates.

The relative timing of each process is chosen randomly and is changed after each update round such that we are sampling equally among all possibilities (see Materials and Methods). This approach reflects the lack of experimental data on relative reaction speeds. The internal states of signaling proteins and the concentrations of small molecules are not explicitly known for each stomate, and components such as $\text{Ca}^{2+}$ and cell membrane potential show various states even in a homogenous experimental setup [249, 250]. Accordingly, we sample a large number (10,000) of randomly selected initial states for the nodes other than ABA and closure (closure is initially set to 0), and let the system evolve either with ABA always on (1) or ABA always off (0). We quantify the probability of closure (equivalent to the percentage of closed stomata in the population) by the formula:

$$P(closure)^t = \frac{1}{N} \sum_{j=1}^{N} S_{\text{closure}}^t(j)$$

where $S_{\text{closure}}^t(j)$ is the state of the node “Closure” at time $t$ in the $j$th simulation and $N$ is the total number of simulations, in our case 10,000. We illustrate the main steps of our simulation method in Figure 3.4.
Figure 3.4. Schematic Illustration of Our Modeling Methodology and of the Probability of Closure

In this four-node network example, node A is the input (as ABA is the input of the ABA signal transduction network), and node D is the output (corresponding to the node “Closure” in the ABA signal transduction network). The nodes’ states are indicated by the shading of their symbols: open symbols represent the off (0) state and filled symbols signify the on (1) state. To indicate the connection between this example and ABA-induced closure, we associate D = off (0) with a picture of an open stomate, and D = on (1) with a picture of a closed stomate. The Boolean transfer functions of this network are $A^* = 1$, $B^* = A$, $C^* = A$, $D^* = B$ and C (i.e., node A is on commencing immediately after the initial condition, the next states of nodes B and C are determined by A, and D is on only when both B and C are on).

(A) The first column represents the networks’ initial states; the input and output are not on, but some of the components in the network are randomly activated (e.g., middle row, node B). The input node A turns on right after initialization, signifying the initiation of the ABA signal. The next three columns in (A) represent the network’s intermediary states during a sequential update of the nodes B, C, and D, where the updated node is given as a gray label above the gray arrow corresponding to the state transition. This sequence of three transitions represents a round of updates from timestep 1 (second column) to timestep 2 (last column). Out of a total of $2^2 \times 3! = 24$ possible different normal responses, two sketches of normal responses are shown in the top two rows. The bottom row illustrates a case in which one node (shown as a square) is disrupted (knocked out) and cannot be regulated or regulate downstream nodes (indicated as dashed edges).

(B) The probability of closure indicates the fraction of simulations where the output $D = 1$ is reached in each timestep; thus, in this illustration the probability of closure for the normal response (circles) increases from 0% at time step 1 to 100% at timestep 2. The knockout mutant’s probability of closure (squares) is 0% at both time steps.
As shown in Figure 3.5, in eight steps, the system shows complete closure in response to
ABA. In contrast, without ABA, although some initial states lead to closure at the beginning,
within six steps the probability of closure approaches 0. Initial theoretical analysis of the
attractors (stable behaviors) of this nonlinear dynamic system confirms that when given a
constant ABA = 1 input, the majority of nodes will approach a steady state value within three to
eight steps. This steady-state value does not depend on the initial conditions. For example, OST1,
PLC, and InsPK stabilize in the on state, and PEPC settles into the off state within the first
timestep when ABA is consistently on. The exception is a set of 12 nodes, including Ca²⁺, Ca²⁺
ATPase, NO, K⁺ efflux from the vacuole to the cytosol, and K⁺ efflux through rapidly activating
K⁺ channels (AP channels) at the plasma membrane (KAP), whose attractors are limit cycles
(oscillations) according to the model. Ca²⁺ oscillations have indeed been observed
experimentally [251, 252]; no time course measurements have been reported in the literature for
the other components, so it is unknown whether they oscillate or not. We identified four subsets
of behaviors for these nodes—distinguished by different positions on the limit cycle—depending
on the initial conditions and relative process durations. Due to the functional redundancy between
K⁺ efflux mechanisms driving stomatal closure (see last entry of Table 3.1), and the stabilization
of the other regulators of the node “Closure” a closed steady state (Closure=1) is attained within
eight steps for any initial condition. The details of this analysis will be published elsewhere.

3.3.4 Identification of Essential Components

After testing the wild-type (intact) system, we investigate whether the disruption (loss) of
a component changes the system’s response to ABA. We systematically perturb the system by
setting the state of a node to 0 (off state), and holding it at 0 for the duration of the simulation.
This perturbation mimics the effect of a knockout mutation for a gene or pharmaceutical inhibition of secondary messenger production or of kinase or phosphatase activity. We characterize the effect of the node disruption by calculating the percentage (probability) of closure response to a constant ABA signal at each time step and comparing it with the percentage of closure in the wild-type system.

The perturbed system’s responses can be classified into five categories with respect to the system’s steady state and the time it takes to reach the steady state. We designate responses identical or very close to the wild-type response as having normal sensitivity; in these cases the probability of closure reaches 100% within eight timesteps. Disruptions that cause the percentage of closed stomata to decrease to zero after the first few steps are denoted as conferring ABA insensitivity (in accord with experimental nomenclature). We observe responses where the probability of closure (the percentage of stomata closed at any given timestep) settles at a nonzero value that is less than 100%; we classify these responses as having reduced sensitivity. Finally, in two classes of behavior the probability of closure ultimately reaches 100%, but with a different timing than the normal response. We refer to a response with ABA-induced closure that is slower than wildtype as hyposensitivity, while hypersensitivity corresponds to ABA-induced closure that is faster than wild-type. Therefore, the perturbed system’s responses can be classified into five categories in the order of decreasing sensitivity defect: insensitivity to ABA, reduced sensitivity, hyposensitivity, normal sensitivity, and hypersensitivity.

We find that 25 single node disruptions (65%; compare with Table 3.2) do not lead to qualitative effects: 100% of the population responds to ABA with timecourses very close to the wild-type response. In contrast, the loss of membrane depolarizability, the disruption of anion efflux, and the loss of actin cytoskeleton reorganization present clear vulnerabilities: irrespective
of initial conditions or of relative timing, all simulated stomata become insensitive to ABA 
(Figure 3.5A). Indeed, membrane depolarization is a necessary condition of K\(^+\) efflux, which is a 
necessary condition of closure, as is actin cytoskeleton reorganization and anion efflux. The 
individual disruption of seven other components — PLD, PA, SphK, S1P, GPA1, K\(^+\) efflux 
through slowly activating K\(^+\) channels at the plasma membrane (KOUT), and pH\(_c\) increase — 
reduces ABA sensitivity, as the percentage of closed stomata in the population decreases to 
20%—80% (see Figure 3.5B). At least five components (S1P, SphK, PLD, PA, pH\(_c\)) of these 7 
predicted components have been shown to impair ABA induced closure when clamped or 
mutated experimentally [25, 148, 165, 253]. For these disruptions, both theoretical analysis and 
numerical results indicate that all simulated stomata converge to limit cycles (oscillations) driven 
by the Ca\(^{2+}\) oscillations, yet the ratio of open and closed stomata in the population is the same at 
any timepoint, leading to a constant probability of closure. (The alternative possibility, of a subset 
of stomata being stably closed and another subset stably open, was not observed for any 
disruption.)

<table>
<thead>
<tr>
<th>Number of Nodes Disrupted</th>
<th>Percentage with Normal Sensitivity</th>
<th>Percentage Causing Insensitivity</th>
<th>Percentage Causing Reduced Sensitivity</th>
<th>Percentage Causing Hyposensitivity</th>
<th>Percentage Causing Hypersensitivity</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>65%</td>
<td>7.5%</td>
<td>17.5%</td>
<td>5%</td>
<td>5%</td>
</tr>
<tr>
<td>2</td>
<td>38%</td>
<td>16%</td>
<td>27%</td>
<td>12%</td>
<td>6%</td>
</tr>
<tr>
<td>3</td>
<td>23%</td>
<td>25%</td>
<td>31%</td>
<td>13%</td>
<td>7%</td>
</tr>
</tbody>
</table>

In all the perturbations, there are five groups of responses. Normal sensitivity refers to a response 
close to the wild-type response (shown as black triangles and dashed line in Figure 3.5). 
Insensitivity means that the probability of closure is zero after the first three steps (see Figure 
5A). Reduced sensitivity means that the probability of closure is less than 100% (see dashed 
symbols in Figure 3.5B). Hyposensitivity corresponds to ABA-induced closure that is slower 
than wild-type (black diamonds in Figure 3.5C). Hypersensitivity corresponds to ABA-induced 
closure that is faster than wild-type (black squares in Figure 3.5C).

For all other single-node disruptions the probability of closure ultimately reaches 100%
(i.e., all simulated stomata reach the closed steady state); however, the rate of convergence diverges from the rate of the wild-type response (see Figure 3.5C). Disruption of Ca\(^{2+}\) increase or of the production of ROS leads to ABA hyposensitivity (slower than wild-type response). In contrast, the disruption of ABI1 or of the Ca\(^{2+}\) ATPase(s) leads to ABA hypersensitivity (faster than wild-type response) (Figure 3.5C). The hyposensitive and hypersensitive responses are statistically distinguishable (p < 0.05 for all intermediary time steps [i.e., for 0 < t < 8]) from the normal responses. Our model predicts that perturbation of OST1 leads to a slower than normal response that is nevertheless not slow enough to be classified as hyposensitive. Indeed, ost1 mutants are still responsive to ABA even though not as strongly as wild-type plants [226].
Figure 3.5. The Probability of ABA-Induced Closure (i.e., the Percentage of Simulations that Attain Closure) as a Function of Timesteps in the Dynamic Model

In all panels, black triangles with dashed lines represent the normal (wildtype) response to ABA stimulus. Open triangles with dashed lines show that in wild-type, the probability of closure decays in the absence of ABA.

(A) Perturbations in depolarization (open diamonds) or anion efflux at the plasma membrane (open squares) cause total loss of ABA-induced closure. The effect of disrupting actin reorganization (not shown) is identical to the effect of blocking anion efflux.

(B) Perturbations in S1P (dashed squares), PA (dashed circles), or pHc (dashed diamonds) lead to reduced closure probability. The effect of disrupting SphK is nearly identical to the effect of disrupting S1P (dashed squares); perturbations in GPA1 and PLD, KOUT are very close to perturbations in PA (dashed circles); for clarity, these curves are not shown in the plot.

(C) abi1 recessive mutants (black squares) show faster than wild-type ABA-induced closure (ABA hypersensitivity). The effect of blocking Ca\(^{2+}\) ATPase(s) (not shown) is very similar to the effect of the abi1 mutation. Blocking Ca\(^{2+}\) increase (black diamonds) causes slower than wild-type ABA-induced closure (ABA hyposensitivity). The effect of disrupting atroboh or ROS production (not shown) is very similar to the effect of blocking Ca\(^{2+}\) increase.
After analyzing all single knockout simulations, we turned to analysis of double and triple knockout simulations. First, to effectively distinguish between normal, hypo- and hypersensitive responses (all of which achieve 100% probability of closure, but at different rates), we calculated the cumulative percentage of closure (CPC) by adding the probability of closure over 12 steps; the smaller the CPC value, the more slowly the probability of closure reaches 100%, and vice versa. Plotting the histogram of CPC values reveals a clear separation into three distinct groups of response in the case of single disruptions (Figure 3.6A). In contrast, the cumulative effects of multiple perturbations lead to a continuous distribution of sensitivities in a broad range around the normal (Figure 3.6B and 3.6C). We use the single perturbation results to identify three classes of response that achieve 100% closure, but at varying rates. We define two CPC thresholds: the midpoint between the most hyposensitive single mutant and normal response, CPC_{hypo} = 10.35; and the midpoint between the normal and least hypersensitive single mutant response, CPC_{hyper} = 10.7. Disruptions with cumulative closure probability < CPC_{hypo} are classified as hyposensitive, disruptions with cumulative closure probability > CPC_{hyper} are hypersensitive; and values between the two thresholds are classified as normal responses. This hypo/hypersensitive classification does not affect the determination of insensitive or reduced sensitivity responses, which are identified by observing a null or less than 100% probability of closure.
Figure 3.6. Classification of Close-to-Normal Responses

(A) For all the single mutants that ultimately reach 100% closure, we plot the histogram of the cumulative probability of closure (CPC). We find three distinct types of responses: hypersensitivity (CPC > 10.7, for \( abi1 \) and \( \text{Ca}^{2+}\text{ATPase} \) disruption); hyposensitivity (CPC < 10.35, for \( \text{Ca}^{2+} \), \( \text{atrboh} \), and ROS disruption); and normal responses (10.35 < CPC < 10.7). For all the double (B) and triple (C) mutants that eventually reach 100% closure at steady state when ABA = 1, we classify the responses using the CPC thresholds defined by the single mutant responses. The CPC threshold values are indicated by dashed vertical lines in the plot.

For double (triple) knockout simulations, some combinations of perturbations exhibit sensitivities that are independent of the sensitivity of each of their components’ perturbation. Normal ABA-induced stomatal closure is preserved in 38% (23%) of combinations (see Table 3.2). In contrast, ABA signaling is completely blocked in 16% (25%) of disruptions. In addition to perturbations involving the three previously found insensitivity-causing single knockouts (loss of membrane depolarizability, the disruption of anion efflux, and the loss of actin cytoskeleton...
reorganization), a large number of novel combinations are found. Interestingly, perturbations of Ca\(^{2+}\), or Ca\(^{2+}\) release from stores, when combined with disruptions in PLD, PA, GPA1, or pH, lead to insensitivity (see **Figure 3.7** and Discussion). ABA-induced closure is reduced (but not lost entirely) in 27% (31%) of the cases. Hyposensitive responses are found for 12% (13%) of double (triple) perturbations. All of the double perturbations in this category involve a knockout mutation of Ca\(^{2+}\), Atrboh, or ROS. The triple perturbations involve a knockout mutation of Ca\(^{2+}\), Atrboh, or ROS, plus two other perturbations, or combinations of three disruptions that alone are not predicted to cause quantifiable effects (e.g., guanyl cyclase, Ca\(^{2+}\) release from internal stores [CIS], and CaIM; see Figure 3.7). Around 6% (7%) of double (triple) perturbations, all including a knockout mutation of ABI1 or Ca\(^{2+}\) ATPase, lead to a hypersensitive response. In summary, accumulating perturbations cause a dramatic decrease in the percentage of normal response; the majority of triple knockouts are either insensitive or have reduced sensitivity. The fraction of hyposensitive and hypersensitive knockouts increases only moderately.
Figure 3.7. Summary of the Dynamic Effects of Calcium Disruptions
All curves represent the probability of ABA-induced closure (i.e., the percentage of simulations that attain closure) as a function of time steps. Black triangles with dashed line represent the normal (wild-type) response to ABA stimulus; open triangles with dashed lines show how the probability of closure decays in the absence of ABA. CIS + PA double mutants (dashed circles) and Ca$_{c}^{2+}$ + pH$_{c}$ double mutants (dashed diamonds) show insensitivity to ABA. Ca$_{c}^{2+}$ ATPase + RCN1 double mutants (black circles) show hyposensitive (delayed) response to ABA. Guanyl cyclase + CIS + CalM triple mutants (black diamonds) also show hyposensitivity; note that none of the guanyl cyclase or CIS or CalM single knockouts show changed sensitivity (data not shown). Ca$_{c}^{2+}$ ATPase mutants (black squares) show faster than wild-type ABA-induced closure (ABA hypersensitivity).

3.3.5 Experimental Assessment of Model Predictions

As a first step toward experimental assessment of the model’s predictions, we used a weak acid, Na-butyrate, to clamp cytosolic pH, and then we treated the stomata with 50 µM ABA and observed the stomatal aperture responses. As shown in Figure 3.8A, the stomatal aperture distributions without butyrate treatments shift towards smaller apertures after ABA treatment, forming a distribution that overlaps with, but is clearly distinguishable from, the 0 ABA distribution. However, when increasing concentrations of butyrate are added in the solution, the “open” (0 ABA) and “closed” (+ ABA) distributions become increasingly overlapping (Figure 3.8B – 3.8D). At the highest butyrate concentration (5 mM; Figure 3.8D), the 0 ABA and +ABA populations of stomatal apertures are statistically identical (the null hypothesis that the two
distributions are the same cannot be rejected; two-tailed t test, p > 0.05). These results qualitatively support our prediction of the importance of pHc signaling.

![Figure 3.8. Effect of Cytosolic pH Clamp (Increasing Concentrations of Nabutyrate from 0 to 5 mM) on ABA-Induced Stomatal Closure](image)

The histograms show the distribution of stomatal apertures without ABA treatment (gray bars) and with 50 µM ABA (white bars). Throughout, the x-axis gives the stomatal aperture size and the y-axis indicates the fraction of stomata for which that aperture size was observed. The black columns indicate the overlap between the 0 µM ABA and the 50 µM ABA distributions. Note that the data of (A) and those of Figure 3.3A are identical; these data are reproduced here for ease of comparison with panels (B–D).

For a more quantitative comparison with the theoretically predicted probability of closure corresponding to pH clamping, one can define a threshold C between open and closed stomatal states, such that stomata with apertures larger than C can be classified as open and stomata with lower apertures can be classified as closed. We identify the threshold value C = 4.3 µm by
simultaneously minimizing the fraction of stomata classified as closed in the control condition and maximizing this fraction in the ABA treated condition. Using this threshold we find that the fraction of closed stomata in the 50 µM ABA + 5 mM Na-butyrate population is 26%, in agreement with the theoretically predicted probability of closure (Figure 3.5B).

In plant systems, cytosolic pH changes in response to multiple hormones such as ABA [231, 254], jasmonates [72], auxin [254], etc. The downstream effectors of pH changes include ion channels [25], protein kinases [255], and protein phosphatases [146]. Previous experiments with guard cells have demonstrated the efficacy of butyrate in imposing a cytosolic pH clamp [25, 72]. While these prior experiments focused on a single concentration of butyrate, here we used five different concentrations (three shown), with 120 stomata sampled for each treatment. As seen in Figure 3.8, we were able to monitor the effect of butyrate in the +ABA treatment in both increasing the mean aperture size and reducing the spread of the aperture sizes. There is a clear indication of saturation between the two highest butyrate concentrations. While detailed measurements of cytosolic pH constitute a full separate study beyond the scope of the present article, the results of Figure 3.8 support the suggestion from our model that pHc should receive increased attention by experimentalists as a focal point for transduction of the ABA signal.
3.4 Discussion

3.4.1 Network Synthesis and Path Analysis

Logical organization of large-scale data sets is an important challenge in systems biology; our model provides such organization for one guard cell signaling system. As summarized in Appendix A. Table S1, we have organized and formalized the large amount of information that has been gathered on ABA induction of stomatal closure from individual experiments. This information has been used to reconstruct the ABA signaling network (Figure 3.2). Figure 3.2 uses different types of edges (lines) to depict activation and inhibition, and also uses different edge colors to indicate whether the information was derived from our model species, *Arabidopsis*, or from another plant species. Different types of nodes (metabolic enzymes, signaling proteins, transporters, and small molecules) are also color coded. An advantage of our method of network construction over other methods such as those used in Science’s Signal Transduction Knowledge Environment (STKE) connection maps [256] is the inclusion of intermediate nodes when direct physical interactions between two components have not been demonstrated.

As is evident from Figure 3.2, network synthesis organizes complex information sets in a form such that the collective components and their relationships are readily accessible. From such analysis, new relationships are implied and new predictions can be made that would be difficult to derive from less formal analysis. For example, building the network allows one to “see” inferred edges that are not evident from the disparate literature reports. One example is the path from S1P to ABI1 through PLD. Separate literature reports indicate that PLDa null mutants show increased transpiration, that PLDa1 physically interacts with GPA1, that S1P promotion of stomatal closure is reduced in *gpa1* mutants, that PLD catalyses the production of PA, and that recessive *abi1*
mutants are hypersensitive to ABA. Network inference allows one to represent all this information as the S1P → GPA1 → PLD → PA → ABI1 → closure path, and make the prediction that ABA inhibition of ABI1 phosphatase activity will be impaired in sphingosine kinase mutants unable to produce S1P.

Another prediction that can be derived from our network analysis is a remarkable redundancy of ABA signaling, as there are eight paths that emanate from ABA in Figure 3.2 and, based on current knowledge (though see below) these paths are initially independent. The prediction of redundancy is consistent with previous, less formal analyses [257]. The integrated guard cell signal transduction network (which includes the ABA signal transduction network) has been proposed as an example of a robust scale-free network [257]. To classify a network as scale-free, one needs to determine the degree (the number of edges, representing interactions/ regulatory relationships) of each node, and to calculate the distribution of node degrees (denoted degree distribution) [8, 258]. Scale-free networks, characterized by a degree distribution described by a power law, retain their connectivity in the face of random node disruptions, but break down when the highest-degree nodes (the so-called hubs) are lost [258]. While the guard cell network may ultimately prove to be scale-free, the network is not sufficiently large at present to verify the existence of a power-law degree distribution; thus, the analogy with scale-free networks cannot be rigorously satisfied.
3.4.2 Dynamic Modeling

Our model differs from previous models employed in the life sciences in the following fundamental aspects. First, we have reconstructed the signaling network from inferred indirect relationships and pathways as opposed to direct interactions; in graph theoretical terminology, we found the minimal network consistent with a set of reachability relationships. This network predicts the existence of numerous additional signal mediators (intermediary nodes), all of which could be targets of regulation. Second, the network obtained is significantly more complex than those usually modeled in a dynamic fashion. We bridge the incompleteness of regulatory knowledge and the absence of quantitative dose-response relationships for the vast majority of the interactions in the network by employing qualitative and stochastic dynamic modeling previously applied only in the context of gene regulatory networks [248].

Mathematical models of stomatal behavior in response to environmental change have been studied for decades [259, 260]. However, no mathematical model has been formulated that integrates the multitude of recent experimental findings concerning the molecular signaling network of guard cells. Boolean modeling has been used to describe aspects of plant development such as specification of floral organs [261], and there are a handful of reports describing Boolean models of light and pathogen-, and light by carbon-regulated gene expression [262-264]. Use of a qualitative modeling framework for signaling networks is justified by the observation that signaling networks maintain their function even when faced with fluctuations in components and reaction rates [265]. Our model uses experimental evidence concerning the effects of gene knockouts and pharmacological interventions for inferring the downstream targets of the corresponding gene products and the sign of the regulatory effect on these targets. However, use of this information does not guarantee that the dynamic model will reproduce the dynamic
outcome of the knockout or intervention. Indeed, all model ingredients (node states, transfer functions) refer to the node (component) level, and there is no explicit control over pathway-level effects. Moreover, the combinatorial transfer functions we employed are, to varying extents, conjectures, informed by the best available experimental information (see Appendix A. Text S1). Finally, in the absence of detailed knowledge of the timing of each process and of the baseline (resting) activity of each component, we deliberately sample timescales and initial conditions randomly. Thus, an agreement between experimental and theoretical results of node disruptions is not inherent, and would provide a validation of the model.

The accuracy of our model is indeed supported by its congruency with experimental observation at multiple levels. At the pathway level, our model captures, for example, the inhibition of ABA-induced ROS production in both ost1 mutants and atrboh mutants [72, 99, 226] and the block of ABA-induced stomatal closure in a dominant-positive atrac1 mutant [176]. In our model, as in experiments, ABA-induced NO production is abolished in either nos single or nia12 double mutants [61, 230]. Moreover, the model reproduces the outcome that ABA can induce cytosolic K⁺ decrease by K⁺ efflux through the alternative potassium channel KAP, even when ABA-induced NO production leads to the inhibition of the outwardly-rectifying (KOUT) channel [109]. At the level of whole stomatal physiology, our model captures the findings that anion efflux [239, 266] and actin cytoskeleton reorganization [176] are essential to ABA-induced stomatal closure. The importance of other components such as PA, PLD, S1P, GPA1, KOUT, pH, in stomatal closure control [25, 84, 148, 165, 231, 253], and the ABA hypersensitivity conferred by elimination of signaling through ABI1 [144, 145], are also reproduced. Our model is also consistent with the observation that transgenic plants with low PLC expression still display ABA sensitivity [267].

The fact that our model accords well with experimental results suggests that the inferences and assumptions made are correct overall, and enables us to use the model to make
predictions about situations that have yet to be put to experimental test. For example, the model predicts that disruption of all Ca\(^{2+}\) ATPases will cause increased ABA sensitivity, a phenomenon difficult to address experimentally due to the large family of calcium ATPases expressed in Arabidopsis guard cells (unpublished data). Most of the multiple perturbation results presented in Figure 3.5 and Table 3.2 also represent predictions, as very few of them have been tested experimentally. Results from our model can now be used by experimentalists to prioritize which of the multitude of possible double and triple knockout combinations should be studied first in wet bench experiments.

Most importantly, our model makes novel predictions concerning the relative importance of certain regulatory elements. We predict three essential components whose elimination completely blocks ABA-induced stomatal closure: membrane depolarization, anion efflux, and actin cytoskeleton reorganization. Seven components are predicted to dramatically affect the extent and stability of ABA-induced stomatal closure: pHc control, PLD, PA, SphK, S1P, G protein signaling (GPA1), and K\(^+\) efflux. Five additional components, namely increase of cytosolic Ca\(^{2+}\), Atrboh, ROS, the Ca\(^{2+}\) ATPase(s), and ABI1, are predicted to affect the speed of ABA-induced stomatal closure. Note that a change in stomatal response rate may have significant repercussions, as some stimuli to which guard cells respond fluctuate on the order of seconds [268, 269]. Thus our model predicts two qualitatively different realizations of a partial response to ABA: fluctuations in individual responses (leading to a reduced steady-state sensitivity at the population level), and delayed response. These predictions provide targets on which further experimental analysis should focus.

Six of the 13 key positive regulators, namely increase of cytosolic Ca\(^{2+}\), depolarization, elevation of pHc, ROS, anion efflux, and K\(^+\) efflux through outwardly rectifying K\(^+\) channels, can be considered as network hubs [8], as they are in the set of ten highest degree (most interactive) nodes. Other nodes whose disruption leads to reduced ABA sensitivity, namely SphK, S1P,
GPA1, PLD, and PA, are part of the ABA → PA path. While they are not highly connected themselves, their disruption leads to upregulation of the inhibitor ABI1, thus decreasing the efficiency of ABA-induced stomatal closure. Similarly, the node representing actin reorganization has a low degree. Thus the intuitive prediction, suggested by studies in yeast gene knockouts [9, 270], that there would be a consistent positive correlation between a node’s degree and its dynamic importance, is not supported here, providing another example of how dynamic modeling can reveal insights difficult to achieve by less formal methods. This lack of correlation has also been found in the context of other complex networks [271].

Comparing Figure 3.3 and Figure 3.6C, one can notice a similar heterogeneity in the measured stomatal aperture size distributions and the theoretical distribution of the cumulative probability of closure in the case of multiple node disruptions. While apparently unconnected, there is a link between the two types of heterogeneity. Due to stochastic effects on gene and protein expression, it is possible that in a real environment not all components of the ABA signal transduction network are fully functional. Therefore, even genetically identical populations of guard cells may be heterogeneous at the regulatory and functional level, and may respond to ABA in slightly different ways. In this case, the heterogeneity in double and triple disruption simulations provides an explanation for the observed heterogeneity in the experimentally normal response: the latter is actually a mixture of responses from genetically highly similar but functionally nonidentical guard cells.

3.4.3 Importance of Ca^{2+} Oscillations to ABA-Induced Stomatal Closure

Through the inclusion of the nodes CaIM, CIS, and the Ca^{2+} ATPase node representing the Ca^{2+} ATPases and Ca^{2+}/H^{+} antiporters [272, 273] that drive Ca^{2+} efflux from the cytosolic compartment, our model incorporates the phenomenon of oscillations in cytosolic Ca^{2+}
concentration, which has been frequently observed in experimental studies [110, 251, 274]. In experiments where \( \text{Ca}^{2+}_c \) is manipulated, imposed \( \text{Ca}^{2+}_c \) oscillations with a long periodicity (e.g., 10 min of \( \text{Ca}^{2+}_c \) elevation with a periodicity of once every 20 min) are effective in triggering and maintaining stomatal closure, yet at 10 min (i.e., after just one \( \text{Ca}^{2+}_c \) transient elevation and thus before the periodicity of the \( \text{Ca}^{2+}_c \) change can be “known” by the cell), significant stomatal closure has already occurred [251]. This result suggests that the \( \text{Ca}^{2+}_c \) oscillation signature may be more important for the maintenance of closure than for the induction of closure [110, 251], and that the induction of closure might only be dependent on the first, transient \( \text{Ca}^{2+}_c \) elevation.

According to our model, if \( \text{Ca}^{2+}_c \) elevation occurs, then stomatal closure is triggered (consistent with numerous experimental studies), but \( \text{Ca}^{2+}_c \) elevation is not required for ABA-induced stomatal closure. Re-evaluation of the experimental studies on ABA and \( \text{Ca}^{2+}_c \) reveals support for this prediction. First, although \( \text{Ca}^{2+}_c \) elevation certainly can be observed in guard cell responses to ABA, numerous experimental results also show that \( \text{Ca}^{2+}_c \) elevation is only observed in a fraction of the guard cells assayed [22, 224]. Furthermore, absence of \( \text{Ca}^{2+}_c \) elevation in response to ABA does not prevent the occurrence of downstream events such as ion channel regulation [275, 276] and stomatal closure [277, 278], a phenomenon also predicted by our in silico analysis. Second, it has been observed that some guard cells exhibit spontaneous oscillations in \( \text{Ca}^{2+}_c \), and in such cells, ABA application actually suppresses further \( \text{Ca}^{2+}_c \) elevation [250]; thus, ABA and \( \text{Ca}^{2+}_c \) elevation are clearly decoupled.

Our model does predict that disruption of \( \text{Ca}^{2+}_c \) signaling leads to ABA hyposensitivity, or a slower than normal response to ABA. In the real-world environment, even a slight delay or change in responsiveness may have significant repercussions, as some stimuli to which guard cells respond fluctuate on the order of seconds; and stomatal responses can have comparable rapidity [268, 269]. Moreover, our model predicts that \( \text{Ca}^{2+}_c \) elevation (although not necessarily oscillation) becomes required for engendering stomatal closure when pH change, \( \text{K}^+ \) efflux or
the S1P-PA pathway are perturbed (see Figure 3.7). Thus, Ca\textsuperscript{2+} modulation confers an essential redundancy to the network. Support for such a redundant role can be found in a study by Webb et al. [279] where Ca\textsuperscript{2+} concentration was reduced below normal resting levels by intracellular application of BAPTA (such reduction in baseline Ca\textsuperscript{2+} levels has been shown to reduce ABA activation of anion channels [276]) and the epidermal tissue was perfused with CO2-free air, a treatment that has been shown to inhibit outwardly rectifying K\textsuperscript{+} channels and slow anion efflux channels [280]. The ABA insensitivity of stomatal closure found by Webb et al. under these conditions [279] therefore can be attributed to a combination of multiple perturbations (of Ca\textsuperscript{2+} elevation, K\textsuperscript{+} efflux, and anion efflux) and is consistent with the predictions of our model.

Our model indicates that double perturbations of the Ca\textsuperscript{2+} ATPase component and either of RCN1, OST1, NO, NOS, NIA12, or Atrboh are hyposensitive (see Figure 3.7), consistent with experimental results on disruptions in the latter components [28, 61, 72, 99, 226, 230]. Since the latter disruptions alone, with unperturbed Ca\textsuperscript{2+} ATPase, are found to have a close-to-normal response in our model, a Ca\textsuperscript{2+} ATPase - disrupted and therefore Ca\textsuperscript{2+} oscillation - free model seems to be closer to experimental observations on stomatal aperture response recorded for these individual mutant genotypes. This suggests that Ca\textsuperscript{2+} elevation (and not Ca\textsuperscript{2+} oscillation) is the signal perceived by downstream factors that control the induction of closure. Possibly, certain as-yet undiscovered interaction motifs, such as a synergistic feedforward loop [281] or dual positive feedback loops [282], could transform the Ca\textsuperscript{2+} oscillation into a stable downstream output.

3.4.4 Limitations of the Current Analysis

**Network topology.** Our graph reconstruction is incomplete, as new signaling molecules will certainly be discovered. Novel nodes may give identity to the intermediary nodes that our model currently incorporates. Discovery of a new interaction among known nodes could simplify
the graph by reducing (apparent) redundancy. For example, if it is found that GPA1 \(\rightarrow\) OST1, the simplest interpretation of the ABA \(\rightarrow\) ROS pathway becomes ABA \(\rightarrow\) GPA1 \(\rightarrow\) OST1 \(\rightarrow\) ROS, and the graph loses one edge and an alternative pathway. As an effect, the graph’s robustness will be attenuated. Among likely candidates for network reduction are the components currently situated immediately downstream of ABA because, in the absence of information about guard cell ABA receptors [283], we assumed that ABA independently regulates eight components. It is also possible that a newly found interaction will not change the existing edges, but only add a new edge. A newly added positive regulation edge will further increase the redundancy of signaling and correspondingly its robustness. Newly added inhibitory edges could possibly damage the network’s robustness if they affect the main positive regulators of the network, especially anion channels and membrane depolarization. For example, experimental evidence indicates that \textit{abi1} \textit{abi2} double recessive mutants are more sensitive to ABA-induced stomatal closure than \textit{abi1} or \textit{abi2} single recessive mutants [145], suggesting that ABI1 and ABI2 act synergistically. Due to limited experimental evidence, we do not explicitly incorporate ABI2, but an independent inhibitory effect of ABI2 would diminish ABA signaling.

While it is difficult to estimate the changes in our conclusions due to future knowledge gain, we can gauge the robustness of our results by randomly deleting entries in Appendix A. Table S1 or rewiring edges of Figure 3.2 (see Appendix A. Texts S2 and S3). We find that most of the predicted important nodes are documented in more than one entry, and more than one entry needs to be removed from the database before the topology of the network related to that node changes (Appendix A. Text S2). Random rewiring of up to four edge pairs shows that the dynamics of our current network is moderately resilient to minor topology changes (Appendix A. Text S3 and Appendix A. Figure S1).

**Dynamic model.** In our dynamic model we do not place restrictions on the relative timing of individual interactions but sample all possible updates randomly. This approach reflects
our lack of knowledge concerning the relative reaction speeds as well as possible environmental
noise. The significance of our current results is the prediction that whatever the timing is, given
the current topology of regulatory relationships in the network, the most essential regulators will
not change. Our approach can be iteratively refined when experimental results on the strength and
timing of individual interactions become available. For example, we can combine Boolean
regulation with continuous synthesis and degradation of small molecules or signal transduction
proteins [284, 285] as kinetic (rate) data emerge. Our model considers the response of individual
guard cell pairs to the local ABA signal; however, there is recent evidence of a synchronized
oscillatory behavior of stomatal apertures over spatially extended patches in response to a
decrease in humidity [286]. Our model can be extended to incorporate cell-to-cell signaling and
spatial aspects by including extracellular regulators when information about them becomes
available (see [246]).

**Node disruptions.** A knockout may either deprive the system of an essential signaling
element (the gene itself), or it may “set” the entire system into a different state (e.g., by affecting
the baseline expression of other, seemingly unrelated signaling elements). Our analysis and
current experimental data only address the former. Because of this caveat, in some ways rapid
pharmacological inhibition may actually have a more specific effect on the cell than gene
knockouts.

### 3.4.5 Implications

Many of the signaling proteins present as nodes in our model are represented by
multigene families in *Arabidopsis* [1], with likely functional redundancy among encoded
isoforms. Therefore, the amount of experimental work required to completely disrupt a given
node may be considerable. It is also considerable work to make such genetic modification in
many of the important crop species that are much less amenable than *Arabidopsis* to genetic manipulation. It is also the case that, at present, there are no reports of successful use of ratiometric pH indicators in the small guard cells of *Arabidopsis*, suggesting that further technical advances in this area are required. Facts such as these indicate the importance of establishing a prioritization of node disruption in experimental studies seeking to manipulate stomatal responses for either an increase in basic knowledge or an improvement in crop water use efficiency. Our model provides information on which such prioritization can be based. Future work on this model will focus on predicting the changes in ABA-induced closure upon constitutive activation of network components or in the face of fluctuating ABA signals. Ultimately, the experimental information obtained may or may not support the model predictions; the latter instance provides new information that can be used to improve the model. Through such iteration of in silico and wet bench approaches, a more complete understanding of complex signaling cascades can be obtained.

Approaches to describe the dynamics of biological networks include differential equations based on mass-action kinetics for the production and decay of all components [287, 288], and stochastic models that address the deviations from population homogeneity by transforming reaction rates into probabilities and concentrations into numbers of molecules [289]. The great complexity of many cellular signal transduction networks makes it a daunting task to reconstruct all the reactions and regulatory interactions in such explicit biochemical and kinetic detail. Our work offers a roadmap for synthesizing incompletely described signal transduction and regulatory networks utilizing network theory and qualitative stochastic dynamic modeling. In addition to being the practical choice, qualitative dynamic descriptions are well suited for networks that need to function robustly despite changes in external and internal parameters. Indeed, several analyses found that the dynamics of network motifs crucial for the stable dynamics and noiseresistance of cellular networks, such as single input modules, feed-forward
loops [290, 291] and dual positive feedback loops [282], is correctly and completely captured by qualitative modeling [292, 293]. For example, at the regulatory module level, several qualitative (Boolean and continuous/discrete hybrid) models [246, 248, 285] reproduced the Drosophila segment polarity gene network’s resilience when facing variations in kinetic parameters [245], offering the most natural explanation of which parameter sets will succeed in forming the correct gene expression pattern [294]. We expect that our methods will find extensive applications in systems where modeling is currently not possible by traditional approaches and that they will act as a scaffold on which more quantitative analyses of guard cell signaling in particular and cell signaling in general can later be built.

Our analyses have clear implications for the design of future wet bench experiments investigating the signaling network of guard cells and for the translation of experimental results on model species such as Arabidopsis to the improvement of water use efficiency and drought tolerance in crop species [295-297]. Drought stress currently provides one of the greatest limitations to crop productivity worldwide [298, 299], and this issue is of even more concern given current trends in global climate change [300, 301]. Our methods also have implications in biomedical sciences. The use of systems modeling tools in designing new drugs that overcome the limitation of traditional medicine has been suggested in the recent literature [302]. Many human diseases, such as breast cancer [303] or acute myeloid leukemia [304, 305], cause complex alterations to the underlying signal transduction networks. Pathway information relevant to human disease etiologies has been accumulated over decades and such information is stored in several databases such as TRANSPATH [306], BioCarta (http://www.biocarta.com), and STKE (http://www.stke.org). Our strategy can serve as a tool that guides experiments by integrating qualitative data, building systems models, and identifying potential drug targets.
3.6 Materials and Methods

**Plant material and growth conditions.** Wild-type *Arabidopsis* (Col genotype) seeds were germinated on 0.5 × MS media plates containing 1% sucrose. Seedlings were grown vertically under short-day conditions (8 h light/16 h dark) 120 µmol m⁻² s⁻¹ for 10 d. Vigorous seedlings were selected for transplantation into soil and were grown to 5 wk of age (from germination) under short day conditions (8 h light/16 h dark). Leaves were harvested 30 min. after the lights were turned on in the growth chamber.

**Stomatal aperture measurements.** Leaves were incubated in 20 mM KCl, 5 mM Mes-KOH, and 1 mM CaCl₂ (pH 6.15) (Tris), at room temperature and kept in the light (250 µmol m⁻² s⁻¹) for 2 h to open stomata. For pHc clamping, different amounts of Na-butyrate stock solution (made up as 1M solution in water [pH 6.1]) were added into the incubation solution, to achieve the concentrations given in Figure 8, 15 min before adding 50 µM ABA. Apertures were recorded after 2.5 h of further incubation in light. Epidermal peels were prepared at the end of each treatment. The maximum width of each stomatal pore was measured under a microscope fitted with an ocular micrometer. Data were collected from 40 stomata for each treatment and each experiment was repeated three times.

**Model.** The network in Figure 3.2 was drawn with the SmartDraw software (http://www.smartdraw.com/exp/ste/home). The dynamic modeling was implemented by custom Python code (http://www.python.org). To equally sample the space of all possible timescales, the random-order asynchronous updating method developed in [248] was used. Briefly, every node is updated exactly once during each unit time interval, according to a given order. This order is a permutation of the N = 40 nodes in the network, chosen randomly out of a uniform distribution over the set of all N! possible permutations. A new update order is selected at each timestep. As demonstrated in [248], this algorithm is equivalent to a random timing of each node’s state.
transition.
Chapter 4 High kurtosis is an essential principle of gene expression in multicellular organisms

4.1 Abstract

Understanding tissue-related gene expression patterns can provide important insights into the organizing principles underlying multicellularity. By analyzing thousands of single gene expression distributions in multiple tissues of human, mouse, Arabidopsis, and rice, we found that a highly leptokurtic distribution of expression levels consisting of a narrow central peak with a heavier than normal rightward tail is a universal attribute of differential gene expression in these multicellular organisms. For human and Arabidopsis, we further used this attribute to identify gene-tissue associations and generate gene-centric tissue networks. The two networks show strikingly similar topological features and predictive power for tissue- and developmentally-related gene functions. We validated our computational predictions by functional analysis: four leptokurtic genes newly predicted to be associated with plant guard cells were found to function in light and abscisic acid responses, and in development of this cell type. Multi-tissue association and multi-level expression patterns underlie the observed leptokurtic distributions and represent a high dimension extension to the genome expansion theory of the evolution of multicellular organisms. This chapter is a manuscript in preparation.
4.2 Introduction

Understanding patterns of gene expression at the global transcriptome level can reveal fundamental principles that govern the complex biological systems of multicellular organisms. Genome-wide tissue-specific gene expression data have been generated for several multicellular organisms [307-310], and gene co-expression has been widely used to identify gene functions in these organisms. For example, if co-expression of a gene pair and its orthologs is conserved across two or more organisms, their biological function also tends to be conserved (Stuart et al., 2003). The co-expression relationships between many pairs of genes compose a gene co-expression network, in which hub genes are thought to be essential and conserved across organisms, and genes in the same network module are likely to have related function [13]. These properties of gene co-expression patterns have been used to prioritize disease candidate genes in humans [15, 311] and to identify new genes involved in known metabolic pathways of plants [312, 313].

However, a co-expression network generated from gene expression data of many tissue types does not directly address a unique feature of multicellular organisms: many genes only play important functions in a subset of all tissue types. If a pair of genes only has important function in some tissues, one would not expect this pair of genes to be co-expressed across all other tissues. In multicellular organisms, gene expression is differentially regulated, giving rise to “tissue-specific” gene expression [314] which further leads to tissue-specific cellular networks [315]. Computational approaches, such as outlier detection [316, 317], multiple comparisons [318, 319], standardized Z-score [310], and clustering analysis [308, 320], have been applied to identify tissue specific genes. However, different methods with different criteria for tissue specificity select different genes: some methods define tissue specific genes as genes that are only expressed
in single tissue type [316], others set limits on the number of tissues in which a “tissue specific”
gene can be expressed [318].

Multicellularity emerged independently several times over the course of evolutionary
history [321]. Despite all the efforts in identifying tissue specific genes, an essential question has
yet to be answered: is there a unified principle of gene expression patterns across distantly related
multicellular organisms? To gain a fundamental understanding of how genes are expressed in
multicellular organisms, we analyzed gene expression from human, mouse, rice and Arabidopsis,
the multicellular organisms for which the most extensive tissue-based gene expression datasets
are available. Because most of these datasets contain gene expression data from organs, tissues
and cell types, for simplicity, we will usually refer to these sample sources as “tissues”. We
started our analysis by asking a basic question that is not addressed by co-expression and tissue-
specific analysis methods: for each single gene, what are the distributional features of gene
expression across multiple tissues? Expression distributions from all genes in a single tissue have
been shown to have similar patterns across many organisms [322]; however, despite the large
amounts of expression data available, little attention has been paid to the shape of the distribution
for each individual gene of an organism across multiple tissues.

To study the shape of a gene expression distribution, one needs to generate a histogram
from the expression levels of that gene in multiple tissue types (Figure 4.1A). By looking at the
expression histogram, one can immediately observe whether the gene shows tissue preferential
expression, how many tissues this gene is preferentially expressed in, and whether this gene is
expressed at the same level or different levels in those preferentially expressed tissues. To
quantify the shape of gene expression distributions for thousands of genes, we calculated
expression kurtosis for every single gene across dozens of tissue types in human, mouse,
Arabidopsis and rice. A high kurtosis or “leptokurtic” expression distribution has a sharp peak
with heavy tails on one or both sides of the peak and the kurtosis of that distribution is larger than
three, which is the kurtosis of a Gaussian distribution. We found that leptokurtic expression of a gene across multiple tissue types is a unifying principle of tissue preferential expression in these four multicellular organisms. Out of the three theoretically possible leptokurtic patterns (heavy tail on the right, on the left, or on both sides of the sharp peak), leptokurtic gene expression overwhelmingly has a specific conserved pattern consisting of a sharp peak with a rightward heavy tail.
**Figure 4.1.** Distribution of gene expression kurtosis in four multicellular organisms.

A. Cartoon of gene assessment strategy. **a)** Gene 1 is highly expressed in only one tissue type. Gene 2 is highly expressed in two tissue types, and expression levels in the two tissues are different. Gene 3 shows no obvious tissue association. **b)** The density of expression levels of gene 1 has one large peak and one small peak (high kurtosis). The density of expression levels of gene 2 has three peaks (moderate kurtosis). The density of expression levels of gene 3 has only one major peak (kurtosis is approximately three). **c)** Genes are ranked based on their expression kurtosis from greatest to least. Gene 1 and gene 2 have larger kurtosis than gene 3. An FDR-based threshold on kurtosis can be used to discriminate gene 1 and gene 2 from gene 3.

B. Cumulative distribution of gene expression kurtosis for all genes from human, mouse, Arabidopsis and rice datasets. **Solid curve:** cumulative distribution of expression kurtosis. Y axis shows the percentage of genes with kurtosis above the values indicated on the X axis (kurtosis). **Dashed curve:** cumulative distribution of expression kurtosis simulated from random backgrounds generated using the same sample size, median, and median-absolute-deviation as the actual microarray datasets. Vertical line indicates the kurtosis threshold corresponding to a false discovery rate (FDR) threshold (Q) of 0.005. This threshold occurs at Human $K = 5.39$, Arabidopsis $K = 5.40$, Mouse $K = 5.59$, Rice $K = 6.55$.

C. Leptokurtic genes ($Q<0.005$) have multi-tissue expression patterns.

Numbers of human and Arabidopsis leptokurtic genes associated with one or more tissue types are plotted as bar graphs. Genes with $Q < 0.005$ were divided into 6 bins based on their expression kurtosis. Total number of genes in each bin was plotted. Black regions indicate number of genes associated with only one tissue type and open regions indicate number of genes associated with more than one tissue type. Because black regions are barely discernible, the percentage of genes that are associated with more than one tissue type is given above each bar.
4.3 Results

4.3.1 Kurtosis of gene expression patterns in four multicellular organisms

We analyzed gene expression data obtained under normal physiological conditions [307, 308, 310, 320]; Materials and Methods and Appendix B.1). For each single gene in each of the four organisms, we first calculated kurtosis of the gene across all tissues. The reason for choosing kurtosis over other metrics to measure tissue preferential expression is discussed in Appendix B.2. We ranked genes based on their expression kurtosis from the greatest to the least (Figure 4.1A), and applied the Anscombe test [323] to identify genes that have excessive kurtosis (as measured by FDR adjusted p value (Q) < 0.005; Materials and Methods and Figures 4.1A, 4.1B). We find 78.8% of human genes, 79.9% of mouse genes, 62.1% of Arabidopsis genes and 39.6% of rice genes have leptokurtic distributions. Most genes are leptokurtic due to expression at higher levels in a small number of tissues; only a few leptokurtic genes have lower expression or have both higher and lower expression in a small number of tissues.

To confirm our analysis, we compared the cumulative distributions of expression kurtosis from all genes within each organism (Figure 4.1B, solid curves) to random backgrounds (Figure 4.1B, dashed curves), which were generated from normal distributions using parameters (sample number, median, median-absolute-deviation) obtained from the original gene expression data for all genes in that organism (Appendix B.2). On average, less than 0.2% of genes from random backgrounds are above our kurtosis threshold. These results confirm that, in all four organisms, many more genes have higher kurtosis than is seen for random backgrounds. Further analysis
focused on the human and Arabidopsis datasets as representative organisms for animals and plants; mouse and rice data were not analyzed further.

4.3.2 Gene tissue associations and gene tissue network analysis

For each leptokurtic gene in the human and Arabidopsis datasets, we asked in which tissues the expression of the gene appears in the upper-tail of the leptokurtic distribution, i.e. in which tissues is the gene preferentially expressed. If we exclude the expression levels in tissues in which the leptokurtic gene is preferentially expressed, the distribution of the leptokurtic gene will then have lower kurtosis. Accordingly, we plotted reduction of kurtosis (ROK) for all genes that passed the kurtosis test in both human and Arabidopsis at several different relative expression thresholds (Z), and chose a threshold in Z such that more than 99% of genes will have a ROK when expression levels above this threshold are excluded from the calculation of kurtosis (Appendix B.3, Appendix B. Figure S1). Interestingly, we found that the ROK curve is very similar in both human and Arabidopsis, hence, we used the same threshold in Z (Z = 3, for Q < 0.005) for both organisms.

We asked whether leptokurtic genes tend to be single-tissue-specific genes, which are the only type of tissue specific genes identified by some previous methods [316]. We divided leptokurtic genes into 6 groups based on their kurtosis, and then, for each group, plotted the number of leptokurtic genes associated with single or multiple tissue types. In both human and Arabidopsis, the majority of leptokurtic genes are associated with more than one tissue type (Figure 4.1C). We further found that almost all high kurtosis genes (HKGs, Q < 1e-9)) exhibit multi-tissue association and a multi-level hierarchy of gene expression levels (Appendix B.4 and Appendix B. Figure S2)
The observation of multi-tissue association of leptokurtic genes enables us to understand the functional relationships between tissues from a fundamentally different perspective than co-expression analysis. Because each leptokurtic gene can be associated with multiple tissues, and each tissue can be associated with many leptokurtic genes, we are able to assign tissue similarities by the relative numbers of leptokurtic genes that are associated with each pair of tissues. To illustrate these tissue similarities as well as the relative numbers of leptokurtic genes that are associated with each tissue, we used HKGs to construct gene-centric tissue networks (GCTN) for both human and Arabidopsis (Figures 4.2, 4.3, Appendix B. Tables S1 and S2, Materials and Methods, and Appendix B.5). Gene-centric tissue networks use a list of leptokurtic genes as signature for each tissue type; this is different from co-expression networks where the expression levels of genes are used to calculate tissue similarities. A higher Q (Q < 1e-09) was used to make visualization of the GCTN less crowded, and the corresponding Z thresholds were determined anew using the ROK method (Appendix B.3, Appendix B. Figure S1). We confirmed the quality of our gene-tissue associations by comparison to manually annotated tissue-related expression designations from the UniProt knowledgebase [324]. Gene-tissue associations defined by our method agree better with these manually annotated expression patterns than gene-tissue specificities defined using other methods [307, 308, 315] (Appendix B.6, Appendix B. Figure S3). Searchable versions of our networks are available at http://personal.psu.edu/szl116/TissueNetworks.

In both human and Arabidopsis datasets, sometimes a whole organ and a single cell type within that organ both appeared in the network. This sample nesting could contribute to the observed multi-level expression patterns and some connections in the GCTNs. However, we confirmed that many leptokurtic genes have multi-level expressions in completely non-
overlapping tissues. In fact, 82% and 73% of connections in the human GCTN (Figure 4.2) and Arabidopsis GCTN (Figure 4.3), respectively, are not due to tissue nesting (Appendix B.7); therefore our subsequent analyses are not confounded by tissue nesting effects.
Figure 4.2. Human gene-centric tissue network.

Human gene-centric tissue network (GCTN), obtained using 1e-09 as the FDR threshold of kurtosis (Q < 1e-09), 10 as the relative expression level threshold (Z = 10), and 20 as the tissue similarity score threshold (S = 20). These more stringent thresholds were used to reduce the number of less significant connections between tissues so that the modularity of the graph can be easily visualized.

Circle diameters are proportional to the number of high kurtosis genes (HKGs) associated with each tissue. Edge colors indicate the tissue similarity score (S), i.e. the significance score of gene overlaps between tissue types. Eight modules were obtained using a weighted label passing algorithm (Appendix B.9). Module numbers are arbitrary. Node labels and detailed annotations can be found in Appendix B. Table S1.
Figure 4.3. Arabidopsis tissue network.

Arabidopsis GCTN, obtained using the same parameters as for the human network (Q < 1e-09, Z = 10, S = 20). Module 1: roots. Module 2 seeds. Module 3, seedling green parts. Module 4, stems. Module 5, young leaves and vegetative shoot apex. Module 6, rosette leaves, senescing leaves and cauline leaves. Module 7, shoot apex, Module 8, floral organs and siliques. Node labels and detailed annotations can be found in Appendix B. Table S2.
In the GCTNs, nodes represent tissues and the size of each node is proportional to the number of HKGs associated with that node. Tissue pairs in the networks are connected by an edge if they share a significant number of commonly associated HKGs, with significance defined by tissue similarity scores (S) (Materials and Methods). We tested whether the commonly used co-expression method [325] generates similar networks for human and Arabidopsis (Appendix B, Table S1, S2, Appendix B.5). We found that 66% of human GCTN edges and 75% of Arabidopsis GCTN edges are different from those found in the corresponding co-expression network, suggesting that GCTNs and co-expression networks capture different properties of the expression data (see Discussion).

We observe a surprising congruency in the GCTNs of these two highly dissimilar multicellular organisms (Figures 4.2-4.4). First, both networks have similar average degree: the human network has 72 nodes and 443 edges, with an average degree (number of edges per node) of 6.2, while the Arabidopsis network has 73 nodes and 476 edges, with an average degree of 6.5 (both networks are obtained with Q = 1e-09, Z = 10 and S = 20). Other graph theoretical properties are also similar in the human and Arabidopsis GCTNs (Figures 4.4, Appendix B. Figure S4 and Appendix B.8). Each species’s network has one giant connected component over a wide range of S (from 1 to 20 for Q = 1e-9 and Z = 10), with a majority of the nodes able to reach the other nodes in the same network (Figure 4.4A). The total numbers of edges in both networks decrease with similar patterns as S increases (Figure 4.4B). Figure 4.4C and 4.4D show the numbers of edges and numbers of nodes in the giant connected components of both tissue networks. In both organisms, these numbers show a sudden drop at around S = 30, which suggests the breakdowns of tissue networks into disconnected networks of smaller sizes.
Second, both human and plant networks show clear modularity, i.e. more connections between tissues within a group than connections between tissues in different groups. This property was revealed by applying a weighted label propagating algorithm for module identification (Appendix B.9). Seven (Figure 4.2) and eight (Figure 4.3) modules were identified for human and Arabidopsis, respectively. Furthermore, we found that each computationally derived module actually corresponds to a group of tissue and cell types with related biological functions, and the numbers of genes belonging to the modules have consistent patterns under different thresholds of \( Q \) and \( Z \) (Figure 4.5 and Appendix B.10). For the Arabidopsis GCTN, seeds (Module 5) always have the highest number of leptokurtic genes, while leaves (Module 7) have the lowest number of leptokurtic genes. For the human GCTN, blood cells (Module 6)

**Figure 4.4.** Graph theoretical properties of tissue networks with \( Q = 1e-09 \) and \( Z = 10 \) and different \( S \) values.

Circles: human tissue network. Triangles: Arabidopsis tissue network. Four graph theoretical properties are plotted with different thresholds of tissue similarity score (\( S \)): A. Number of connected components. B. Total number of edges in the network. C. Number of edges in the giant connected component. D. Total number of nodes in the giant connected component. For all panels, x axis represents tissue similarity score (\( S \)).
always have the highest number of leptokurtic genes, while ganglions (Module 1) have the lowest number of leptokurtic genes. These patterns suggest that some tissues tend to use more leptokurtic genes to perform their physiological function than other tissues. Within a given module, tissues with similar physiological roles have higher S than tissues that are less similar. For example, CD4 T cells, CD8 T cells and early erythroid cells all belong to the blood cells group, but CD4 T cells and CD8 T cells have an S of 320, whereas CD4 T cells and early erythroid cells only have an S of 21 (Figure 4.2 module 6; Appendix B.11 for more examples from Arabidopsis).

**Figure 4.5.** Panel A: Median numbers of leptokurtic genes in each module of the human tissue network. The number one threshold corresponds to Q = 0.005 and Z = 3; the number eight threshold corresponds to Q = 1e-09 and Z = 10. Other thresholds are linear interpolations of these two sets of Q and Z values. See Appendix A.10 for details. **Panel B:** Median numbers of leptokurtic genes in each module of the Arabidopsis tissue network. The number one threshold in both figures corresponds to Q = 0.005 and Z = 3; the number eight threshold corresponds to Q = 1e-09 and Z = 10. Other thresholds are linear interpolations of these two sets of Q and Z values.
Third, we discovered that the S scores of both organisms coincide with developmental hierarchies. For example, connections between fetal liver and liver, and between fetal lung and lung, are stronger than connections with other tissues in the network and in the same module (Figure 4.2 module 7). Developmental hierarchies are more obviously seen in the Arabidopsis network, because more samples are from developing tissues in the Arabidopsis dataset than in the human dataset. For example, in the subnetwork containing seeds and siliques, or in the subnetwork containing shoot apical meristems and early flowers, the S scores are higher between tissues that are in consecutive developmental stages than between those that are not (Figure 4.6A and B). For genes that appear in the nodes of these two subnetworks, we further studied gene function using Gene Ontology (GO) (Appendix B. Figure S5 and S6, and Appendix B.12), and found that many enriched GO annotations agree well with known functions of the corresponding tissues. For example, in the silique to seed subnetwork, pectinesterase (GO:0030599) and inhibitor of pectinesterase activity (GO: 0046910) are enriched in early silique development, reflecting cell wall restructuring during fruit development, while seed development (GO:0048316), negative regulation of seed germination (GO:0010187) and embryonic development ending in seed dormancy (GO:0009793) become enriched in later stages of seed development (Appendix B. Figure S5). Given that GO functions of leptokurtic genes suffice to reflect known biological functions of these developmental stages, other enriched GO groups (Appendix B. Figures S5 and S6) represent novel discoveries of genes and gene functions that likely also participate in these developmental stages.

Fourth, although nodes are less connected between modules, intermodule links (edges) correspond with many known functional connections between tissue types, and consequently provide information about gene functions. For example, fetal liver and adult liver are both complex organs that play diverse physiological roles, however, one well-known difference
between fetal liver and adult liver is that fetal liver is one of the main sites of haematopoiesis [326], while adult liver does not perform this function. This functional difference is clearly captured in the subnetwork of the unique first tissue neighbors of fetal liver and adult liver (Figure 4.6C). Unique first neighbors of fetal liver are bone marrow, CD34 cells, early erythroid and endothelial cells, which are all related to haematopoiesis, while the unique first neighbors of adult liver are lung, adrenal gland, adipocyte and heart, which are not related to haematopoiesis. GO analysis for the leptokurtic genes associated with fetal liver, adult liver, and their neighboring tissues validates this observation: genes associated with fetal liver and their unique first neighbors are especially enriched in the process of erythrocyte maturation (GO:0043249). These results from GO analysis support that leptokurtic genes play important roles in tissues with which they are associated (Appendix B.13 and Appendix B. Figure S7 for more examples).

Overall, these connections clearly show that tissue networks of leptokurtic genes can capture functional similarities and differences between tissue types, and can rediscover known genes with tissue preferential functions. Many genes identified by our methods currently are annotated as “unknown” or have no annotations related to the functions of the associated tissue. Our analysis suggests that these genes have a high likelihood to play important roles in the tissues with which they are associated, and thus are candidates for targeted wet bench analyses.
Figure 4.6. Panel A: Arabidopsis shoot apex subnetwork and Panel B Arabidopsis seed subnetwork. Diameters of circles are proportional to the numbers of leptokurtic genes in each tissue. Edge colors indicate the tissue similarity score (S). Connections between shoot apex and early floral tissues are plotted. Other connections, such as those between early floral tissues (Stage 9) and late floral tissues (after Stage 12), were not plotted for clarity. Both networks were obtained with $Q < 1e^{-09}$, and $Z = 10$. $S = 2$ was used to include more weak edges. The nodes were organized following on their developmental stages in a clockwise manner. In siliques and seeds subnetwork, Stage 3, Stage 4 and Stage 5 are siliques that contain immature seeds. Stage 6, Stage 7, Stage 8, Stage 9 and Stage 10 are seeds without siliques. Samples from Stage 3 to Stage 10 are in consecutive developmental stages. In shoot apex to early flower subnetwork, Vegetative Young Leaves, Vegetative, Before Bolting and Inflorescence are four samples from consecutive developmental stages of shoot apex, and Stage 9, Stage 10/11 and Stage 12 are flowers at consecutive developmental stages after Inflorescence. We can see that, in the seed sub-network, numbers of genes associated with each tissue are relative constant. In the shoot apex sub-network, as the shoot apex emerges and become early flowers, more and more genes are associated with tissues from the later stages. Also, tissues adjacent to each other in their developmental stages share more genes than tissues across developmental stages. This was not detected by previous analysis of the same expression data using other methods [308].

Panel B: Human liver and fetal liver subnetwork. The common and unique first neighbors of adult liver and fetal liver in the human tissue network are plotted. Color coding of edges and nodes are the same as in Figure 2. Relative sizes of the nodes are the same as in Figure 2. Group 1, Adult liver unique first neighbors. Group 2, Fetal liver unique first neighbors. Group 3, common first neighbors of both adult liver and fetal liver. Group 4, common genes between adult liver and fetal liver.
4.3.3 Functional analysis of leptokurtically distributed genes associated with guard cells

Arabidopsis offers the opportunity for direct wet-bench evaluation of our predictions at the whole organism level using knockout mutants of leptokurtic genes. We focus on Arabidopsis guard cells: guard cell pairs surround stomatal pores and, by reversible changes in cell volume, regulate the apertures of these microscopic pores which are the conduits through which plants take up carbon dioxide for conversion into biomass and also lose water vapor. Guard cells continually sense a wide array of environmental signals and integrate this information into changes in cell solute content that in turn drive the trans-membrane water fluxes that alter cell volume and stomatal pore width. Among these signals, light stimulates stomatal opening and the plant hormone abscisic acid (ABA) inhibits stomatal opening and promotes stomatal closure.

Arabidopsis guard cells have become a model system to study cell signaling mechanisms that are conserved between animal and plants [165] and to test systems biology approaches of elucidating signaling networks [5, 68]. We generated guard cell transcriptome data [327] and used a linear model based approach to combine our dataset with other datasets (Appendix B.14 and 15).

Using the kurtosis-based approach we found that 22 out of 67 known guard cell functional genes [68] that have been identified by forward or reverse genetic approaches (e.g. OST1, OST2, KAT1, KAT2, ABI1, ABI2) are present in the list of 2345 leptokurtic genes associated with guard cells (Appendix B. Table S3). This result confirms that our method can identify genes with recognized function in guard cells. Because GCTN can capture functional connections between tissues, we studied the potential functional connections between guard cells and the first neighbors of guard cells in the GCTN (Table 4.1). Guard cells are connected with root epidermal cells, including root hair cells, non-hair epidermal cells and lateral root caps. These connections are likely due to the fact that guard cells are also a type of epidermal cell. Both
guard cells and other epidermal cells play major roles in plant defense against pathogens and in regulation of water loss. The functional connection between guard cells and epidermal cells is further characterized and confirmed by GO analysis (Appendix B. Table S4). Leptokurtic genes that are associated with guard cells or root epidermal cells are enriched in GO categories of response to chitin (GO:0010200), a well known elicitor for plant defense response [250]. Leptokurtic genes associated with guard cells and two out of the three root epidermal cell types are also enriched in GO categories of response to water deprivation (GO:0009414). We also found that hydrogen peroxide (H$_2$O$_2$) function (GO:0042542) is enriched in leptokurtic genes associated with guard cells and all three root epidermal cells: reactive oxygen species are well documented as secondary messengers in stress responses of both roots and guard cells.

Table 4.1. First neighbor tissues of guard cells.

The Arabidopsis gene-centric tissue network (GCTN) was obtained using Q < 0.005, Z = 3, S = 40. Guard cell neighbor tissue: First neighbors of guard cells. Annotations: Detailed annotations of tissue types from original publications (Brady et al., 2007; Schmid et al., 2005). Tissue similarity scores (S): Calculated as described in the Materials and Methods section. Number of genes: Number of genes associated with each first neighbor tissues and, in parentheses, number of genes shared with guard cells (2345 leptokurtotic genes are associated with guard cells).

<table>
<thead>
<tr>
<th>Guard cell neighbor tissue</th>
<th>Annotations</th>
<th>Tissue similarity score (S)</th>
<th>Number of genes (shared with guard cells)</th>
</tr>
</thead>
<tbody>
<tr>
<td>LRC</td>
<td>Lateral root cap</td>
<td>112</td>
<td>2834 (639)</td>
</tr>
<tr>
<td>S17</td>
<td>Phloem pole pericycle</td>
<td>108</td>
<td>3310 (689)</td>
</tr>
<tr>
<td>scr5</td>
<td>Endodermis</td>
<td>78</td>
<td>2830 (570)</td>
</tr>
<tr>
<td>gl2</td>
<td>Roots, non-hair cells</td>
<td>73</td>
<td>2532 (522)</td>
</tr>
<tr>
<td>ATGE84</td>
<td>Seeds, stage 10</td>
<td>56</td>
<td>3566 (620)</td>
</tr>
<tr>
<td>RM1000</td>
<td>Lateral root primordia initials</td>
<td>56</td>
<td>3006 (539)</td>
</tr>
<tr>
<td>ATGE25</td>
<td>Senescing leaves</td>
<td>52</td>
<td>1975 (400)</td>
</tr>
<tr>
<td>J0571</td>
<td>Ground tissues</td>
<td>50</td>
<td>2417 (455)</td>
</tr>
<tr>
<td>ATGE41</td>
<td>Flowers stage 15, sepals</td>
<td>46</td>
<td>1892 (376)</td>
</tr>
<tr>
<td>ATGE82</td>
<td>Seeds, stage 8</td>
<td>45</td>
<td>2873 (502)</td>
</tr>
<tr>
<td>ATGE83</td>
<td>Seeds, stage 9</td>
<td>42</td>
<td>3525 (572)</td>
</tr>
<tr>
<td>COBL9</td>
<td>Roots, hair cells</td>
<td>42</td>
<td>3224 (524)</td>
</tr>
<tr>
<td>J0121</td>
<td>Xylem pole pericycle</td>
<td>42</td>
<td>2379 (430)</td>
</tr>
<tr>
<td>Wol</td>
<td>Stele</td>
<td>41</td>
<td>2680 (462)</td>
</tr>
</tbody>
</table>
In the first neighbor network of guard cells, we also found connections between guard
cells and three tissues corresponding to seeds at different developmental stages. Both guard cells
and seeds are major target sites of ABA, which regulates both stomatal movement and seed
dormancy/germination. Response to ABA stimulus (GO:0009737) is enriched in guard cells and
in one of the seed nodes. These results also support our conclusion that GCTN can determine
functional connections between tissues. Our analysis finds that many genes are associated with
guard cells and the 14 guard cell neighboring tissues (Table 4.1 and Supporting website).
However, only leptokurtic genes that are associated with guard cells, and not leptokurtic genes
associated with any of the 14 guard cell neighboring tissues, are enriched in the GO:0010118
term “stomatal movement”. This result suggests that our analysis has a high degree of specificity.
The connections seen between guard cells and senescing leaves or flower sepals, and between
guard cells and non-epidermal root cell types including pericycle cells (Table 4.1) are unexpected
based on current knowledge of the physiologies of these cell or tissue types, suggesting that
functional similarities between these cell types exist but have not yet been identified.

Guard cells mainly reside in the epidermis of rosette leaves; however, we found no
connections in our network between guard cells and samples from rosette leaves. As a
comparison, we also analyzed guard cell neighbor tissues identified using the co-expression tissue
network approach [325]. Out of the 14 samples that have largest correlations with guard cells in
co-expression analysis, 12 are samples of rosette leaves at different developmental stages,
suggesting that, in this case, co-expression networks cannot capture intriguing functional
connections that are observed in GCTN.

The functional significance of guard cells’ leptokurtic genes was evaluated
experimentally. Quantitative reverse-transcriptase PCR (Q-PCR) on 7 different Arabidopsis
tissues for 12 leptokurtic genes that are associated with guard cells showed that all 12 genes have
strong preferential expression in guard cells (Appendix B. Figure S8 A), supporting that our
method can correctly identify tissue preferentially expressed genes. The fact that several of the
genes which have relatively lower kurtosis values also have low but non-negligible expression
levels in other tissues tested (Appendix B. Figure S8 A) is consistent with the fact that
leptokurtic gene sets include genes that are not strictly single-tissue specific.

We functionally analyzed four of these genes: At1g50400, a putative mitochondrial
import receptor subunit TOM40 homolog 2/putative voltage-gated ion channel of mitochondria;
At5g60410 (SIZ1) an E3 ubiquitin ligase; At2g21080, annotated as similar to an extracellular
ligand-gated ion channel; and At1g11100, a SNF2 domain-containing and helical domain
containing protein (Figure 4.7). These genes were not selected based on their annotations, but
solely because these genes are leptokurtic genes associated with guard cells and had two
independent T-DNA mutant lines lacking expression of full length transcript available at the time
of this experiment (Appendix B. Figure S8 B and C).
Figure 4.7. Stomatal phenotypes of mutants of four guard cell leptokurtic genes.

ABA-inhibition of light-stimulated stomatal opening and ABA-induction of stomatal closure. Open bars = control, and closed bars = ABA (50 µM) treated stomatal apertures (3 replicates, 300 stomata each, x ± S.E.). Numbers represent percentage change in apertures in the presence of ABA. * represents significant difference from wild-type Col (P<0.01) in response to ABA. ** represents significant difference from wild-type Col (P<0.05) in light-stimulated stomatal opening. Images are from ABA-treated leaves (second alleles are shown in Appendix A. Figure S8 E). Superimposed numbers represent average guard cell length (n = 100 ± S.E.). Scale bar = 10 µm.

T-DNA knockouts of all four genes resulted in altered guard cell physiology as compared with Col wild-type. Two of the mutants (At2g21080, At1g11100) showed enhanced light-stimulation of stomatal opening, while one mutant (At5g60410 (siz1)) showed reduced opening as compared to wild type (Figure 4.7). Three of the mutants (At1g50400, At5g60410, At1g11100) showed hypersensitivity to ABA-inhibition of stomatal opening and ABA-promotion of stomatal closure (Figure 4.7). Morphologically, guard cells of mutants At2g21080 and At1g11100 were significantly longer than wild-type, while siz1 had shorter guard cells and also exhibited a greater percentage of immature guard cells in adult leaves than did wild type (Figure 4.7, Appendix B. Figure S8 D, E).

By GCTN analysis, we found that At5g60410 (SIZ1) is also associated with lateral root caps and ground tissues of roots, which is consistent with the observation that the siz1 mutants are hypersensitive to ABA inhibition of root elongation [328]. Among the mutants characterized, At1g11100 is also associated with both guard cells and root cells in the guard cell GCTN (Table 4.1). We found that At1g11100 mutants are hypersensitive to ABA inhibition of root elongation (Appendix B. Figure S9). Thus, both published results from siz1 mutants and our new experiments on At1g11100 indicate that our method accurately captures the root related functions for both genes inferred from the GCTN. Our GCTN and wet-bench analyses also support the premise that genes important for the specialized function of a cell type/tissue are not necessarily single-tissue specific.
Direct analysis of leptokurtic genes using induced mutation is not possible for human genes. To characterize functions of human genes with leptokurtic expression distributions, we looked at human disease genes as compiled in the Online Mendelian Inheritance In Man (OMIM) database. We found that the leptokurtic gene-set is enriched in human disease genes (Appendix B.16 and Figure 4.8), suggesting that mutation of leptokurtic genes contributes to many diseases processes.

**Figure 4.8.** Disease genes are over-represented in human leptokurtic genes. Human genes were selected using different Q threshold by kurtosis test. The enrichment scores (negative log p value of Fisher’s exact test) of disease genes that map to selected genes were calculated with different Q. For Q = 1e-9 (the peak of the line in this figure), we found that 4149 probesets passed our kurtosis test, and 694 of these probesets map to OMIM (3) genes, representing a significant enrichment of diseases genes (p value < 4.2e-45) in the HKGs.
4.3.4 Orthologs of high kurtosis genes are also high kurtosis genes in the counterpart organism

Because both computational and wet-bench analyses strongly support the importance of genes with leptokurtic expression distributions, we further characterized the general functions of HKGs (Q < 1e-09). We first asked whether these HKGs are likely to be genes that are evolutionarily conserved across the animal and plant kingdoms. To quantify how many HKGs in human are evolutionarily conserved in Arabidopsis and vice-versa, we analyzed groups of evolutionarily conserved genes between human and Arabidopsis from the Inparanoid database [329]. We found under-representation of evolutionarily conserved genes in both human and Arabidopsis HKGs (Fisher exact test, for human, p < 2.2e-16, for Arabidopsis, p < 2.2e-16, Appendix B.17). However, for those HKGs that are conserved, orthologs in the counterpart organism are also likely to be HKGs (both p < 0.001, permutation test, Appendix B.17).

We also identified enriched GO annotations for human and Arabidopsis HKGs. The 10 GO categories that are significantly enriched and also contain the most HKGs show no overlap between the two organisms, except for the GO category “extracellular region” (GO:0005576, Table 4.2, and Appendix B.12). Human HKGs are enriched in annotations such as plasma membrane, organ development and trans-membrane receptor activities, while Arabidopsis HKGs are enriched in annotations such as endomembrane system, transcription factor activity, kinase activity lipid binding, and apoplast.
Table 4.2. GO analysis for human and Arabidopsis HKGs.

Gene ontology enrichment analysis was carried out using the eliminate count method in topGO [330]. **Annotated**: number of genes annotated by this GO term in the relevant genome. **Selected**: number of HKGs annotated by each GO term. This column is called “Significant” by the original topGO software, but is changed here to “Selected” for clarity. **Expected**: expected number of genes annotated by this GO term. **p value**: p value is calculated using the eliminate count method with p value of 0.01 as an eliminating parameter. **Annotation**: a short description of each GO term. All enriched GO terms with p value < 0.01 were obtained and these terms were then ranked from greatest to least based on the “Selected” column. The top 10 GO annotations with the maximum number of “Selected” HKGs among all other significant GO annotations are shown here. This way of ranking GO categories guarantees that generic terms (annotated to many genes) are included such that our interpretation of GO enrichment represents generic biological functions. Only one GO category (extracellular region GO:0005576) is enriched in both human and Arabidopsis HKGs (Q<1e-9).

A. Human top 10 GO annotation

<table>
<thead>
<tr>
<th>GO id</th>
<th>Annotated</th>
<th>Selected</th>
<th>Expected</th>
<th>p value</th>
<th>Annotation</th>
</tr>
</thead>
<tbody>
<tr>
<td>GO:0016021</td>
<td>5072</td>
<td>1150</td>
<td>1025.51</td>
<td>6.58E-03</td>
<td>integral to membrane</td>
</tr>
<tr>
<td>GO:0005886</td>
<td>4201</td>
<td>1029</td>
<td>849.4</td>
<td>1.29E-07</td>
<td>plasma membrane</td>
</tr>
<tr>
<td>GO:0005576</td>
<td>2088</td>
<td>769</td>
<td>422.17</td>
<td>7.18E-41</td>
<td>extracellular region</td>
</tr>
<tr>
<td>GO:0048513</td>
<td>1973</td>
<td>577</td>
<td>401.46</td>
<td>8.16E-03</td>
<td>organ development</td>
</tr>
<tr>
<td>GO:0005887</td>
<td>1785</td>
<td>453</td>
<td>360.91</td>
<td>5.25E-08</td>
<td>integral to plasma membrane</td>
</tr>
<tr>
<td>GO:0008283</td>
<td>1414</td>
<td>359</td>
<td>287.71</td>
<td>2.62E-03</td>
<td>cell proliferation</td>
</tr>
<tr>
<td>GO:0005509</td>
<td>1123</td>
<td>302</td>
<td>225.89</td>
<td>9.09E-09</td>
<td>calcium ion binding</td>
</tr>
<tr>
<td>GO:0006629</td>
<td>1000</td>
<td>298</td>
<td>203.48</td>
<td>8.10E-04</td>
<td>lipid metabolic process</td>
</tr>
<tr>
<td>GO:0004888</td>
<td>1162</td>
<td>297</td>
<td>233.73</td>
<td>1.02E-03</td>
<td>transmembrane receptor activity</td>
</tr>
<tr>
<td>GO:0051239</td>
<td>941</td>
<td>289</td>
<td>191.47</td>
<td>7.33E-03</td>
<td>regulation of multicellular organismal process</td>
</tr>
</tbody>
</table>

B. Arabidopsis top 10 GO annotation

<table>
<thead>
<tr>
<th>GO id</th>
<th>Annotated</th>
<th>Selected</th>
<th>Expected</th>
<th>p value</th>
<th>Annotation</th>
</tr>
</thead>
<tbody>
<tr>
<td>GO:0012505</td>
<td>2617</td>
<td>735</td>
<td>399.03</td>
<td>8.56E-76</td>
<td>endomembrane system</td>
</tr>
<tr>
<td>GO:0003700</td>
<td>1428</td>
<td>257</td>
<td>218.75</td>
<td>2.31E-03</td>
<td>transcription factor activity</td>
</tr>
<tr>
<td>GO:0016301</td>
<td>1236</td>
<td>230</td>
<td>189.34</td>
<td>3.13E-03</td>
<td>kinase activity</td>
</tr>
<tr>
<td>GO:0006468</td>
<td>859</td>
<td>180</td>
<td>129.45</td>
<td>1.41E-06</td>
<td>amino acid phosphorylation</td>
</tr>
<tr>
<td>GO:0005975</td>
<td>739</td>
<td>150</td>
<td>111.37</td>
<td>2.00E-03</td>
<td>carbohydrate metabolic process</td>
</tr>
<tr>
<td>GO:0005576</td>
<td>364</td>
<td>111</td>
<td>55.5</td>
<td>3.70E-10</td>
<td>extracellular region</td>
</tr>
<tr>
<td>GO:0045533</td>
<td>327</td>
<td>98</td>
<td>50.09</td>
<td>1.08E-06</td>
<td>hydrolase activity</td>
</tr>
<tr>
<td>GO:0009535</td>
<td>294</td>
<td>74</td>
<td>44.83</td>
<td>2.61E-04</td>
<td>chloroplast thylakoid membrane</td>
</tr>
<tr>
<td>GO:0048046</td>
<td>271</td>
<td>72</td>
<td>41.32</td>
<td>9.18E-07</td>
<td>apoplast</td>
</tr>
<tr>
<td>GO:0004091</td>
<td>229</td>
<td>70</td>
<td>35.08</td>
<td>3.64E-04</td>
<td>carboxylesterase activity</td>
</tr>
</tbody>
</table>
4.4 Discussion

4.4.1 A new organizing principle of gene expression in multicellular organisms

Traditional analyses of gene expression patterns in multicellular organisms have focused on the identification of co-expressed genes [12, 13], or tissue specific genes [315-318]. Here we show that high kurtosis in gene expression levels is present in four distantly related multicellular organisms. Because these four model animal and plant species have drastically different life styles and because multicellularities of animals and plants seem to be independently evolved [314, 331], we conclude that leptokurtic distribution of gene expression appears to be a novel emergent property in multicellular organisms. We also reveal multi-tissue association and multi-level hierarchy of gene expression levels for HKGs (Figure 4.1 and Appendix B. Figure S2). Such complex expression patterns of HKGs explain why it has been difficult to reach a unified definition of “tissue-specific” in previous studies. Expression distributions from other multicellular organisms such as fruitfly [332] and worm [333] can be evaluated once gene expression data from a sufficient number of tissues (> 40) become available. The sensitivity of our method can be adjusted by using different thresholds in kurtosis and by using the ROK method to find the corresponding expression threshold. Our results demonstrate the importance of kurtosis as a metric for data analysis. In addition, our tissue-network method can be readily applied for analysis and visualization of other high dimension data sources, including quantitative proteomic and metabolomic datasets.
4.4.2 Gene-tissue co-association is complementary to gene co-expression analysis

Traditional co-expression analysis (e.g. the Pearson correlation coefficient) measures expression similarities from all tissues for any pair of genes. Our “co-association” analysis indicates whether any pair of genes is co-activated in any particular tissue type, and therefore is useful to biologists working with specific cell types or tissues. In most other tissue similarity measurements, the expression of a group of selected genes is used to derive all pair-wise tissue similarities [308, 310]; this may be of limited utility in some instances, since genes that are used for co-expression analysis do not necessarily carry tissue specific functions [334, 335]. By contrast, our approach will assign a unique set of leptokurtic genes that defines the similarity of a pair of tissues, thus providing specific predictions as to which genes are likely to function in tissue pairs. Our gene-centric method can capture tissue similarities: tissues within our computationally derived modules have related physiological functions, and the inter- and intra-module connections derived from our method agree well with known functional connections between tissue types. Furthermore, GO functional groups of genes enriched in pairs of connected tissues share expected common functions of these tissue pairs. These results strongly support that our method is able to associate a gene with tissues where the gene has a functional role. Accordingly, our method provides the basis for assigning putative functions to many other unannotated leptokurtic genes based on their tissue association.

In addition to our computational results, the fact that genetic knockout analysis revealed that each of four HKGs of guard cells has clear functional roles in that cell type further supports our method as a valuable tool to predict tissue-related gene functions. In the siz1 and At1g11100 mutants, the observed ABA hypersensitivity of guard cells in our experiments and the ABA hypersensitive root phenotypes [328], support our computational prediction that SIZ1 and At1g11100 are functionally associated with both roots and guard cells. These results together
suggest that our methods provide a systematic way to annotate gene functions by transferring annotation of a gene from one tissue to another. Such predictions are used routinely by experimental biologists on a small scale, e.g. ABA-responsive genes identified by forward genetic screens on seeds have been correctly predicted to also function in ABA responses of guard cells [144] but to our knowledge, have never been described before at a global level.

Our tissue networks also provide the possibility to annotate a group of genes by transferring functions of another group of genes through the commonly associated tissues of both gene groups. For example, response to chitin, response to water deprivation, and hydrogen peroxide function are all enriched in guard cells and three root epidermal cell types. Therefore, for genes that currently lack annotation, but are similarly associated with these four specific epidermal cell types, we can provisionally assign those genes to any of these three enriched GO groups, and then test such hypotheses by further experimentation targeted toward those particular GO functions.

4.4.3 High kurtosis genes and the evolution of multicellularity

We found that genes with high kurtosis are overall not likely to show evolutionary conservation between human and Arabidopsis, suggesting that HKGs either evolved from species-specific ancestor genes or that their orthologs were lost in the counterpart organism during evolution. This result is in agreement with the independent origination model of the evolution of multicellular organisms [336]. However, for those HKGs that are identified in Arabidopsis and have orthologs in human (or vice-versa), the counterpart gene in the other organism is also likely to be a HKG. This suggests that Arabidopsis is a good model system to study the molecular functions of human HKGs [337], and supports previous conclusions that Arabidopsis is a good model system for assessing the functions of some human genes, including
some human disease genes [1, 338]. In addition, the fact that human leptokurtic genes are enriched in known human disease genes suggests that other, as yet uncharacterized leptokurtic genes are also good candidates for functional roles in human tissue-associated diseases.

An additional possibility is that the human and Arabidopsis HKGs are not conserved in their protein sequences, but are functionally conserved. Because animals and plants diverged 1.6 billion years ago, they could have evolved such that proteins with non-homologous sequences nevertheless perform similar functions. According to the genome expansion theory of multicellularity [314, 336], genes with functions of transcriptional regulation, cell-cell communication and cell adhesion were particularly crucial to the evolution of multicellularity, and gene families conferring these functions proliferated in multicellular organisms via gene duplication. While some gene families apparently proliferated in parallel with evolution of multicellularity in animals, there is evidence that genes involved in development of the extracellular matrix and cell-cell adhesion of multicellular animals were already present in pre-animal genomes of unicellular organisms (Rokas, 2008). It is therefore intriguing that GO annotations pertaining to the extracellular region (GO:0005576) are enriched in both human and Arabidopsis HKGs, suggesting that these HKGs may have been important for the evolution of multicellularity in both kingdoms. Similarly, several enriched GO annotations in both organisms are related to cell-cell communication and adhesion, albeit in an organism-dependent manner. For instance, the enriched GO annotation “apoplast” (GO:0048046) corresponds to a plant-specific extracellular matrix and diffusion area outside the plasma membrane that is central to plant-specific cell-cell adhesion and communication. Likewise, in human we find that cell-cell communication-related functions such as transmembrane receptor activity (GO:0004888) are enriched in HKGs. We also find that transcription factor activity (GO:0003700) is enriched in Arabidopsis HKGs but not in human HKGs, likely because the Arabidopsis dataset but not the human dataset has many tissues in consecutive developmental stages. Our enriched GO
annotations further suggest a unique importance of the endomembrane system in plant multicellularity, which may be rooted in the fact that intercellular communication of plants is partly mediated by an endomembrane system that remains connected in mature cells through plasmodesmata. In the future, it would be interesting to assess whether there is evolutionary evidence that other HKGs, in different GO categories, have also contributed to the evolution of multicellularity.

4.4.4 Tissue networks represent higher dimension genome expansion

Genome expansion based on gene family growth explains how genetic materials are “made-ready” for multicellular organisms (Kaiser, 2001; Rokas, 2008). However, genome expansion does not address the question of efficiency: for a gene that is encoded in one fraction of the whole genome (DNA), how is that fraction of genome reused (transcribed) in different tissue types? By network visualization of gene co-activation patterns, our research provides a systematic analysis of such efficiency in two independently originated multicellular organisms. Our analysis shows that human and Arabidopsis GCTNs are similar in their average degree, patterns of node sizes, and sizes and stabilities of giant connected components. These topological similarities suggest that human and Arabidopsis may have evolved to a similar efficiency of gene usage at the transcriptional level.

Besides the extensive co-activations of genes in different tissues, the observed multi-level hierarchy of expression for HKGs (Appendix B.4, Figure 4.1C and Appendix B. Figure S2) suggests another level of complexity in the utilization of genomic materials. Some genes could be essential to the central function of one tissue type, and yet also provide robustness to their associated physiological process via their expression in another cell type. For example, plant leptokurtic genes such as SIZ1 and ABI1 are important for ABA inhibition of stomatal opening in
the leaf yet also regulate root elongation [144, 328]. In plants vulnerable to drought, the expression of ABA response genes such as SIZ1 and ABI1 in guard cells could reduce water loss in a manner complementary to mechanisms available for maintenance of plant water status by regulation of root growth. Thus, a hierarchy of expression levels could contribute to the physiological robustness of multicellular organisms. We also found that the kurtosis of SIZ1 is lower than that of other genes in our wet bench analyses. The lower kurtosis of SIZ1 expression is consistent with the known pleiotropic effects of the siz1 mutation [328].

Our computationally derived tissue modules are seen to consist of tissues with similar physiological functions. We speculate that all leptokurtic genes in a module initially enter a “potent” state such that these genes can be highly activated in the tissues of that module, while these genes are not in a “potent” state in other tissues outside the module. This gives rise to the modular structure. When tissues within a module become more specialized, further modifications of the expression levels of some leptokurtic genes give rise to differences of gene expressions within a module, which leads to the observed multi-level hierarchy of expression. Whether this speculation is correct and whether the proposed “potent” state corresponds to any molecular mechanisms, such as activated chromosome regions [339], are key questions for future research.

In conclusion, our analysis of leptokurtic expression distributions, gene co-activation patterns as illustrated in tissue networks, and multiple hierarchies of expression levels, uncovers rich patterns of gene expression in multicellular organisms. As an addition to the genome expansion theory, which explains the evolution of multicellularity largely by one dimensional (DNA) expansion, i.e. proliferation of gene families, our gene-centric tissue networks reveal the complexity of multicellularity at a high dimensional space.
4.5 Materials and Methods

4.5.1 Publicly available datasets

Gene expression data from 73 human tissues and 61 mouse tissues [307], 78 Arabidopsis tissues [308, 309, 320] and 42 rice tissues [310] were obtained from the public domain (Appendix B.1). Guard cell and whole leaf microarray datasets were generated in our laboratory (GC dataset) [327].

4.5.2 Statistical tests of kurtosis:

Gene expression levels for each gene were centered and scaled (Appendix B.2) before analysis. Anscombe test for kurtosis was carried out using the Anscombe.test function in the R moment package. We define FDR adjusted p value as Q, and use Q < 0.005 as a threshold to select genes with leptokurtic distributions.

4.5.3 Tissue networks

In each GCTN, nodes represent tissue types, with size of the node proportional to the number of leptokurtic genes associated with that tissue (Appendix B.5). For any pair of tissues, a p value of Fisher’s exact test was calculated based on the number of genes associated with each tissue, number of genes associated with both tissues, and the sum of common and unique genes between the two tissues. FDR correction was used on all calculated p values to obtain a Q value. The tissue similarity score (S) is defined as the negative log 10 of the Q value. Two nodes are connected with an edge in Figure 4.2 and 4.3 if S is larger than 20.

4.5.4 Gene ontology analysis

Gene ontology enrichment analysis was carried out using the eliminate count method in topGO software [330]. Details are given in Appendix B.12.

4.5.5 Microarray analysis of guard cell transcriptome
To compare microarray results from different studies, we applied a linear model based strategy to remove systematic differences between microarray expression levels in different studies (Appendix B.15). We validated our own microarray data by quantitative real-time RT-PCR (Q-PCR), and found high consistency between our renormalized results and our Q-PCR data (Appendix B. Figure S8 and Appendix B.14).

4.5.6 T-DNA mutant analysis.

T-DNA insertions were verified using standard techniques. Stomatal aperture assays were conducted similarly to Li et al. (2006) [5] (see Appendix B.14 for details).
Chapter 5 Summary and Future Directions

5.1 Summary

Computational systems biology has become the major discipline for biologists to understand complex biological systems. While we are still far from the complete understanding of how the underlying cellular networks function for most biological systems, the results from this dissertation summarize the author’s contribution to the system level understanding of the guard cell functions, from both bottom-up and top-down perspectives. The approaches discussed here are not limited to the guard cell system but are also applicable to other cell or tissue types.

In chapter 3, we have developed a Boolean network approach to simulate the process of ABA induced stomatal closure. In the Boolean framework, nodes represent signaling proteins, secondary messengers, ion channels, and other components that are involved in the process of ABA induced stomatal closure. The regulatory interactions between these components are represented as directed edges. We assume that each component only has two states: A component that is activated is represented by an “ON” state. A component that is inactivated is represented by an “OFF” state. The change of state of one component is determined by the states of the components that are upstream of this component, and a Boolean rule that describes how upstream components affect the state of the downstream component. The dynamic of each component is then simulated using a random asynchronous update method [5].

Mathematical modeling of cellular signaling networks is a very challenging problem due to the incompleteness of our knowledge of the structure of the networks, the lack of detailed reaction parameters in the networks, and the inherent stochasticity of intracellular signaling processes. Our approach provides a set of tools to address these problems. Our network
reconstruction method incorporates the parsimony principle into the assembly of the signaling network. The Boolean framework used here is essentially parameter free. We also incorporate the uncertainty of cellular signaling networks into the conventional Boolean approach, by using random initial conditions as well as the random asynchronous update algorithm. The simulation results of the model -- namely the probabilities of closure after ABA treatment in both wild type and mutants -- agree well with known responses. Our guard cell model predicts that clamping intracellular pH can diminish ABA induced stomatal closure. This prediction is validated by experiments, supporting that Boolean modeling not only can reproduce known responses, but also can predict responses that were not tested before.

In chapter 4, we have developed a top-down approach, which starts with the discovery that high kurtosis is the unifying principle of tissue preferential expression in four multicellular organisms. The kurtosis based approach is able to accurately associate genes with the tissues in which they are preferentially expressed. The major advantage of the kurtosis based method is that the method does not predetermine the number of tissue a gene “should” be expressed in. From a biological point of view, the number of tissues a gene is preferentially expressed in is determined by the evolution of the regulatory sequences of the gene, the transcription factors that bind to these regulatory sequences and the small RNAs that control transcript levels. As such, one should not expect to find a “threshold” of how many tissues a tissue preferentially expressed gene can be expressed in.

We find that the most stringently defined tissue specific gene: those genes that are only expressed in one tissue type account for only a small minority of tissue preferentially expressed genes. Therefore, the highly restricted expression patterns of single tissue specific genes are not favored by evolution. Instead, many genes in four multicellular organisms are high kurtosis gene. Furthermore, for high kurtosis genes that are preferentially expressed in multiple tissues, we find that most genes are expressed at multiple levels. A gene can act as the major functional
contributor to the function of the tissue in which that gene is expressed at the highest level, while
the same gene could contribute to the secondary functions of other tissues in which the gene is
expressed at a lower level.

Using a network based approach, we study the global structure of how genes are
expressed in multiple tissues, and we found that the tissue network patterns are similar between
human and Arabidopsis, suggesting similar efficiency in gene usage in both organisms. This top-
down approach can also lead to functional characterization of genes. Four genes are shown to
have novel functions in guard cells through the analysis of phenotypes of knockout mutants.

5.2 Future research interests

5.2.1 Extension to the bottom-up modeling approach.

In the ABA induced stomatal closure process, more experimental evidence regarding new
components, such as the ABA receptors GTG1/2 [173] and PYR1[340, 341], and new regulatory
mechanisms, such as the interaction between GTG1/2 and GPA1 and the interaction between
PYR1 and ABI1, have been identified since publication of this model. An immediate extension to
the model is to incorporate these components and interactions into the existing model. The
addition of new components may require addition of new interactions, removal or substitution of
intermediary nodes, as well as changing the Boolean rules in the model. The process of changing
the model to reproduce experimental observations can potentially lead to hypotheses about yet
unobserved regulatory mechanisms in the ABA signaling network.

Another possible extension to the existing ABA induced stomatal closure model is to
formulate the model as a continuous model. Currently, each node in the model can only take two
values, “0” or “1”. Experimentally, one could administrate ABA at different concentrations (from
micromolar to millimolar range) and measure the stomatal aperture at different time points. A continuous model, such as the piece-wise linear method that is implemented in the package BooleanNet [342], has the potential to incorporate these quantitative details.

The work outlined in chapter 3, i.e. the synthesis of signaling network from reported direct and indirect interactions, and the Boolean dynamic modeling approach, can be extended to other signaling networks in Arabidopsis and in other species. One has to first define the system of interest, and then gather the literature information regarding the direct and indirect regulatory processes that are involved in the signaling process. The information can then be synthesized using a software package, Net-Synthesis, which performs the network reconstruction automatically [343]. Then one needs to define the Boolean rules that determine how each node changes its state. Finally, the dynamical simulation now can be carried out using BooleanNet.

5.2.2 Extension to the top-down approach.

The top-down approach has been applied to four different multicellular organisms. However, all data used in this analysis are from microarray experiments. The microarray platforms that are used in these experiments have the limitation that not all genes encoded in the genomes are included. Other technologies, such as tiling arrays [344] and high throughput RNA sequencing (RNA-seq)[345] can potentially measure all protein coding genes as well as genes that encode microRNAs and other small RNAs. Once the multi-tissue expression data become available, we can apply the kurtosis approach to genes that are not sampled by the datasets used in our analysis.

Although more than 70 tissues are used in our analysis for human and Arabidopsis, some tissues, such as Schwann cells in human and trichomes in Arabidopsis, are not sampled in the datasets in our analysis. For tissues that are not included, one can generate gene expression
profiles using the same microarrays (HGU133A for human and ATH1 for Arabidopsis) that are used in our analysis. The linear modeling approach can be used to merge the expression profiles from the newly generated data to the existing dataset. The kurtosis method can then be used to identify gene-tissue associations for the tissues of interest and the GCTN approach can be used to find functional connections between the tissues of interest and other tissues that are already sampled.

The author is currently exploring how to use the GCTN approach to make functional predictions for HKGs. Different tissues in a multicellular organism have specialized functions, while some tissues may also have similar functions. For example, two different tissues could response to one hormone while other tissues cannot response to that hormone. The genes related to the hormone response will express in both hormone-responsive tissues. The idea is to first define functions of tissues according to the known gene functions: for each tissue type, find the enriched GO functional annotations of HKGs in that tissue type. Then functions of a gene can be predicted based on the tissues in which the gene is expressed. One hypothetical scenario is that five tissues out of all 80 tissues in Arabidopsis are enriched in the function of ABA response. Therefore, a gene that is only associated with these five tissues is more likely to be related to ABA response, than genes which are associated with only subset of the five tissues or genes that are associated with tissues other than these five tissues.
Appendix A. Supplemental Information for Chapter 3

A. Effect of Rewiring Positive Interactions

Effect of Rewiring Positive Interactions

Percentage of Each Category

Number of Pairs of Interaction Rewired

B. Effect of Rewiring Negative Interactions

Effect of Rewiring Negative Interactions

Percentage of Each Category

Number of Pairs of Interactions Rewired

- reduced
- reaches zero
- reaches unity
Appendix A. Figure S1: Probability of closure in randomized networks where pairs of positive (A) or negative (B) edges are rewired.

The x axis indicates the number of rewired edge pairs, from zero (control) up to eight. The bar graphs display the distribution of dynamic behaviors in a large number of randomized networks (the sum of the three segments is 100%). Yellow denotes response that reaches unity (100% closure), blue indicates reduced response (less than 100% closure), and maroon means response that reaches zero (0% closure).
Appendix A. Table S1. Synthesis of Experimental Information about Regulatory Interactions between ABA Signal Transduction Pathway Components.

The explanation of the abbreviations used in the first two columns can be found in
the caption of Figure 2 (in the main text). Each entry in the database consists of a source
node (Node A), a target node or target process (Node/Process B), their interaction, the
plant species this interaction was observed in, and references. In the column
“Node/Process B,” → means activation, ---| means inhibition; both can be enzymatic
processes, e.g. protein phosphorylation, protein dephosphorylation etc. In the interaction
column, five different interactions are allowed: “promotes”, “partially promotes”,
“inhibits”, “partially inhibits”, and “no relation”. “A promotes B” is inferred if in the
absence of A, B cannot be observed; “A inhibits B” is inferred if in the absence of A, B is
upregulated. A promotes/inhibits process B means that in the absence of A, process B
cannot happen/is upregulated. “Partially promotes” or “partially inhibits” means
that B is not completely controlled by A. “No relation” means that the node or process B
is not affected by changes in A. The reference list gives representative references; in
some cases the phenomenon was described in multiple reports, and additional literature
references are given in the main text.
<table>
<thead>
<tr>
<th>Node A</th>
<th>Node/Process B</th>
<th>Interaction</th>
<th>Species</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABA</td>
<td>SphK →</td>
<td>Promotes</td>
<td>Arabidopsis</td>
<td>[165]</td>
</tr>
<tr>
<td>ABA</td>
<td>AAPK (OST1 homologue) →</td>
<td>Promotes</td>
<td>Vicia faba</td>
<td>[346]</td>
</tr>
<tr>
<td>ABA</td>
<td>CaM →</td>
<td>Promotes</td>
<td>Vicia faba</td>
<td>[347]</td>
</tr>
<tr>
<td>ABA</td>
<td>NO →</td>
<td>Promotes</td>
<td>Arabidopsis</td>
<td>[61]</td>
</tr>
<tr>
<td>ABA</td>
<td>Ca^{2+}c →</td>
<td>Promotes</td>
<td>Commelina communis</td>
<td>[279]</td>
</tr>
<tr>
<td>ABA</td>
<td>PLD → PA</td>
<td>Promotes</td>
<td>Vicia faba</td>
<td>[253]</td>
</tr>
<tr>
<td>ABA</td>
<td>PLD →</td>
<td>Promotes</td>
<td>Vicia faba</td>
<td>[253]</td>
</tr>
<tr>
<td>ABA</td>
<td>H^+ ATPase →</td>
<td>Partially inhibits</td>
<td>Vicia faba</td>
<td>[74, 348]</td>
</tr>
<tr>
<td>ABA</td>
<td>RAC1 (ROP6) →</td>
<td>Inhibits</td>
<td>Arabidopsis</td>
<td>[176]</td>
</tr>
<tr>
<td>ABA</td>
<td>ADPRc →</td>
<td>Promotes</td>
<td>Arabidopsis</td>
<td>[349]</td>
</tr>
<tr>
<td>ABA</td>
<td>Ca^{2+}c oscillation →</td>
<td>Promotes</td>
<td>Arabidopsis</td>
<td>[237]</td>
</tr>
<tr>
<td>ABA</td>
<td>InsP3 →</td>
<td>Promotes</td>
<td>Vicia faba</td>
<td>[348]</td>
</tr>
<tr>
<td>ABA</td>
<td>InsP6 →</td>
<td>Promotes</td>
<td>S. tuberosum</td>
<td>[350]</td>
</tr>
<tr>
<td>ABA</td>
<td>OST1 →</td>
<td>Promotes</td>
<td>Arabidopsis</td>
<td>[226]</td>
</tr>
<tr>
<td>ABA</td>
<td>ROS →</td>
<td>Partially promotes</td>
<td>Vicia faba</td>
<td>[351]</td>
</tr>
<tr>
<td>ABA</td>
<td>Hyperpolarization → Ca^{2+}c →</td>
<td>Promotes</td>
<td>Vicia faba</td>
<td>[352]</td>
</tr>
<tr>
<td>ABA</td>
<td>PEPC →</td>
<td>Inhibits</td>
<td>Vicia faba</td>
<td>[233]</td>
</tr>
<tr>
<td>ABA</td>
<td>Malate →</td>
<td>Partially inhibits</td>
<td>Commelina communis</td>
<td>[353]</td>
</tr>
<tr>
<td>SphK</td>
<td>ABA → Closure</td>
<td>Partially promotes</td>
<td>Arabidopsis</td>
<td>[165]</td>
</tr>
<tr>
<td>SphK</td>
<td>ABA → AnionEM</td>
<td>Partially promotes</td>
<td>Arabidopsis</td>
<td>[165]</td>
</tr>
<tr>
<td>SphK</td>
<td>ABA → S1P →</td>
<td>Promotes</td>
<td>Arabidopsis</td>
<td>[165]</td>
</tr>
<tr>
<td>S1P</td>
<td>ABA → Closure</td>
<td>Promotes</td>
<td>Arabidopsis</td>
<td>[165]</td>
</tr>
<tr>
<td>GPA1</td>
<td>S1P → AnionEM</td>
<td>Promotes</td>
<td>Arabidopsis</td>
<td>[165]</td>
</tr>
<tr>
<td>GPA1</td>
<td>ABA → ROS →</td>
<td>Promotes</td>
<td>Arabidopsis</td>
<td>Unpublished</td>
</tr>
<tr>
<td>GPA1</td>
<td>ABA → AnionEM</td>
<td>Promotes</td>
<td>Arabidopsis</td>
<td>[25]</td>
</tr>
<tr>
<td>GPA1</td>
<td>PLD →</td>
<td>Promotes</td>
<td>Arabidopsis</td>
<td>[354]</td>
</tr>
<tr>
<td>GCR1</td>
<td>ABA → Closure</td>
<td>Inhibits</td>
<td>Arabidopsis</td>
<td>[355]</td>
</tr>
<tr>
<td>GCR1</td>
<td>ABA → Closure</td>
<td>Partially inhibits</td>
<td>Arabidopsis</td>
<td>[355]</td>
</tr>
<tr>
<td>GCR1</td>
<td>S1P → Closure</td>
<td>Partially inhibits</td>
<td>Arabidopsis</td>
<td>[355]</td>
</tr>
<tr>
<td>PLD</td>
<td>Phospholipids → PA →</td>
<td>Promotes</td>
<td>Vicia faba</td>
<td>[253]</td>
</tr>
<tr>
<td>PLD</td>
<td>ABA → Closure</td>
<td>Promotes</td>
<td>Vicia faba</td>
<td>[253]</td>
</tr>
<tr>
<td>PA</td>
<td>ROS →</td>
<td>Promotes</td>
<td>Arabidopsis</td>
<td>[102]</td>
</tr>
<tr>
<td>PA</td>
<td>ABI1 →</td>
<td>Inhibits</td>
<td>Arabidopsis</td>
<td>[356]</td>
</tr>
<tr>
<td>Protein</td>
<td>Reaction</td>
<td>Effect</td>
<td>Organism</td>
<td>Reference</td>
</tr>
<tr>
<td>---------</td>
<td>----------</td>
<td>--------</td>
<td>----------</td>
<td>-----------</td>
</tr>
<tr>
<td>ROP2</td>
<td>PA → ROS</td>
<td>Partially promotes</td>
<td>Arabidopsis</td>
<td>[102]</td>
</tr>
<tr>
<td>RCN1</td>
<td>ABA → Closure</td>
<td>Promotes</td>
<td>Arabidopsis</td>
<td>[28]</td>
</tr>
<tr>
<td>RCN1</td>
<td>ABA → AnionEM</td>
<td>Promotes</td>
<td>Arabidopsis</td>
<td>[28]</td>
</tr>
<tr>
<td>RCN1</td>
<td>ABA → Ca^{2+}_c</td>
<td>Promotes</td>
<td>Arabidopsis</td>
<td>[28]</td>
</tr>
<tr>
<td>OST1</td>
<td>ABA → Closure</td>
<td>Promotes</td>
<td>Arabidopsis</td>
<td>[226]</td>
</tr>
<tr>
<td>OST1</td>
<td>ABA → ROS</td>
<td>Promotes</td>
<td>Arabidopsis</td>
<td>[226]</td>
</tr>
<tr>
<td>OST1</td>
<td>ROS → Closure</td>
<td>No relation</td>
<td>Arabidopsis</td>
<td>[226]</td>
</tr>
<tr>
<td>PLC</td>
<td>ABA → Ca^{2+}_c</td>
<td>Promotes</td>
<td>Arabidopsis</td>
<td>[250]</td>
</tr>
<tr>
<td>PLC</td>
<td>ABA → Closure</td>
<td>Partially promotes</td>
<td>Vicia faba</td>
<td>[253]</td>
</tr>
<tr>
<td>PLC</td>
<td>ABA → Closure</td>
<td>Promotes</td>
<td>Commelina communis</td>
<td>[267, 357]</td>
</tr>
<tr>
<td>PLC</td>
<td>ABA → Ca^{2+}_c</td>
<td>Partially Promotes</td>
<td>Commelina communis</td>
<td>[357]</td>
</tr>
<tr>
<td>PLC</td>
<td>ABA → Closure</td>
<td>Partially promotes</td>
<td>Nicotiana rustica</td>
<td>[228]</td>
</tr>
<tr>
<td>PLC</td>
<td>ABA → KOUT</td>
<td>Partially promotes</td>
<td>Commelina communis</td>
<td>[358]</td>
</tr>
<tr>
<td>RAC1 (ROP6)</td>
<td>Actin disruption</td>
<td>Partially inhibits</td>
<td>Arabidopsis</td>
<td>[176]</td>
</tr>
<tr>
<td>RAC1 (ROP6)</td>
<td>Closure</td>
<td>Partially inhibits</td>
<td>Arabidopsis</td>
<td>[176]</td>
</tr>
<tr>
<td>RAC1 (ROP6)</td>
<td>ABA → Closure</td>
<td>Partially inhibits</td>
<td>Arabidopsis</td>
<td>[176]</td>
</tr>
<tr>
<td>NIA12</td>
<td>ABA → Closure</td>
<td>Promotes</td>
<td>Arabidopsis</td>
<td>[61]</td>
</tr>
<tr>
<td>NIA12</td>
<td>Nitrite → Closure</td>
<td>Promotes</td>
<td>Arabidopsis</td>
<td>[61]</td>
</tr>
<tr>
<td>NIA12</td>
<td>ABA → NO</td>
<td>Promotes</td>
<td>Arabidopsis</td>
<td>[61]</td>
</tr>
<tr>
<td>NIA12</td>
<td>Nitrite → NO</td>
<td>Promotes</td>
<td>Arabidopsis</td>
<td>[61]</td>
</tr>
<tr>
<td>NIA12</td>
<td>NO → Closure</td>
<td>No relation</td>
<td>Arabidopsis</td>
<td>[61]</td>
</tr>
<tr>
<td>NIA12</td>
<td>ROS → Closure</td>
<td>No relation</td>
<td>Arabidopsis</td>
<td>[61]</td>
</tr>
<tr>
<td>Nitrite</td>
<td>NIA12 → NO</td>
<td>Substrate</td>
<td>Arabidopsis</td>
<td>[61]</td>
</tr>
<tr>
<td>NOS</td>
<td>ABA → Closure</td>
<td>Promotes</td>
<td>Arabidopsis</td>
<td>[230]</td>
</tr>
<tr>
<td>NOS</td>
<td>ABA → NO</td>
<td>Promotes</td>
<td>Arabidopsis</td>
<td>[230]</td>
</tr>
<tr>
<td>NO</td>
<td>ABA → Closure</td>
<td>Promotes</td>
<td>Pisum sativum</td>
<td>[359]</td>
</tr>
<tr>
<td>NO</td>
<td>ABA → Closure</td>
<td>Promotes</td>
<td>Arabidopsis</td>
<td>[61]</td>
</tr>
<tr>
<td>NO</td>
<td>AnionEM</td>
<td>Promotes</td>
<td>Vicia faba</td>
<td>[107]</td>
</tr>
<tr>
<td>NO</td>
<td>ABA → AnionEM</td>
<td>Promotes</td>
<td>Vicia faba</td>
<td>[107]</td>
</tr>
<tr>
<td>NO</td>
<td>KOUT</td>
<td>Partially inhibits</td>
<td>Arabidopsis</td>
<td>[109]</td>
</tr>
<tr>
<td>NO</td>
<td>Ca^{2+}_c → CIS</td>
<td>Partially promotes</td>
<td>Vicia faba</td>
<td>[107]</td>
</tr>
<tr>
<td>NO</td>
<td>Hyperpolarization → CalM</td>
<td>Inhibits</td>
<td>Vicia faba</td>
<td>[107]</td>
</tr>
<tr>
<td>ADPRc</td>
<td>NO → Ca^{2+}_c</td>
<td>Promotes</td>
<td>Vicia faba</td>
<td>[107]</td>
</tr>
<tr>
<td>cADPR</td>
<td>Ca^{2+}_c → CIS</td>
<td>Promotes</td>
<td>Vicia faba</td>
<td>[107]</td>
</tr>
<tr>
<td>Pathway</td>
<td>Reaction</td>
<td>Effect</td>
<td>Species/Plant</td>
<td>Reference</td>
</tr>
<tr>
<td>-----------------</td>
<td>---------------------------</td>
<td>-----------------------------</td>
<td>------------------------</td>
<td>-----------</td>
</tr>
<tr>
<td>cADPR</td>
<td>ABA → KOUT</td>
<td>Partially promotes</td>
<td>Commelina communis</td>
<td>[358]</td>
</tr>
<tr>
<td>GC</td>
<td>NO → Ca&lt;sup&gt;2+&lt;/sup&gt;</td>
<td>Promotes</td>
<td>Vicia faba</td>
<td>[107]</td>
</tr>
<tr>
<td>InsPK</td>
<td>InsP6</td>
<td>Promotes</td>
<td>S. tuberosum</td>
<td>[350]</td>
</tr>
<tr>
<td>InsP6</td>
<td>CIS</td>
<td>Partially promotes</td>
<td>Vicia faba</td>
<td>[229]</td>
</tr>
<tr>
<td>pH&lt;sub&gt;c&lt;/sub&gt;</td>
<td>ABA → AnionEM</td>
<td>Promotes</td>
<td>Arabidopsis</td>
<td>[25]</td>
</tr>
<tr>
<td>pH&lt;sub&gt;c&lt;/sub&gt;</td>
<td>ABA → Closure</td>
<td>Promotes</td>
<td>Arabidopsis</td>
<td>[25]</td>
</tr>
<tr>
<td>pH&lt;sub&gt;c&lt;/sub&gt;</td>
<td>ROS</td>
<td>Partially promotes</td>
<td>Arabidopsis</td>
<td>[72]</td>
</tr>
<tr>
<td>pH&lt;sub&gt;c&lt;/sub&gt;</td>
<td>KOUT</td>
<td>Promotes</td>
<td>Vicia faba</td>
<td>[225]</td>
</tr>
<tr>
<td>pH&lt;sub&gt;c&lt;/sub&gt;</td>
<td>H&lt;sup&gt;+&lt;/sup&gt;ATPase</td>
<td>Inhibits</td>
<td>Vicia faba</td>
<td>[360]</td>
</tr>
<tr>
<td>pH&lt;sub&gt;c&lt;/sub&gt;</td>
<td>Depolarization → KAP</td>
<td>Inhibits</td>
<td>Arabidopsis</td>
<td>[361]</td>
</tr>
<tr>
<td>pH&lt;sub&gt;c&lt;/sub&gt;</td>
<td>ABI1</td>
<td>Promotes</td>
<td>Arabidopsis</td>
<td>[146]</td>
</tr>
<tr>
<td>Atrboh</td>
<td>ABA → CalM</td>
<td>Promotes</td>
<td>Arabidopsis</td>
<td>[99]</td>
</tr>
<tr>
<td>Atrboh</td>
<td>ROS → CalM</td>
<td>No relation</td>
<td>Arabidopsis</td>
<td>[99]</td>
</tr>
<tr>
<td>NADPH</td>
<td>ABA → CalM</td>
<td>Promotes</td>
<td>Arabidopsis</td>
<td>[100]</td>
</tr>
<tr>
<td>NADH</td>
<td>ABA → CalM</td>
<td>Partially promotes</td>
<td>Arabidopsis</td>
<td>[100]</td>
</tr>
<tr>
<td>ROS</td>
<td>ABA → Closure</td>
<td>Partially promotes</td>
<td>Vicia faba</td>
<td>[351]</td>
</tr>
<tr>
<td>ROS</td>
<td>Closure</td>
<td>Partially promotes</td>
<td>Vicia faba</td>
<td>[351]</td>
</tr>
<tr>
<td>ROS</td>
<td>ABA → CalM</td>
<td>Promotes</td>
<td>Vicia faba</td>
<td>[362]</td>
</tr>
<tr>
<td>ROS</td>
<td>KOUT</td>
<td>Inhibits</td>
<td>Vicia faba</td>
<td>[362]</td>
</tr>
<tr>
<td>ROS</td>
<td>ABA ---</td>
<td>H&lt;sup&gt;+&lt;/sup&gt;ATPase</td>
<td>Partially promotes</td>
<td>Vicia faba</td>
</tr>
<tr>
<td>ROS</td>
<td>ABI1</td>
<td>Inhibits</td>
<td>Arabidopsis</td>
<td>[104]</td>
</tr>
<tr>
<td>ERA1</td>
<td>ABA → CalM</td>
<td>Inhibits</td>
<td>Arabidopsis</td>
<td>[181]</td>
</tr>
<tr>
<td>ERA1</td>
<td>ABA → Closure</td>
<td>Inhibits</td>
<td>Arabidopsis</td>
<td>[181]</td>
</tr>
<tr>
<td>ERA1</td>
<td>ABA → CalM</td>
<td>Inhibits</td>
<td>Arabidopsis</td>
<td>[181]</td>
</tr>
<tr>
<td>ERA1</td>
<td>Ca&lt;sup&gt;2+&lt;/sup&gt; → AnionEM</td>
<td>No relation</td>
<td>Arabidopsis</td>
<td>[181]</td>
</tr>
<tr>
<td>ROP10</td>
<td>ABA → Closure</td>
<td>Partially inhibits</td>
<td>Arabidopsis</td>
<td>[177]</td>
</tr>
<tr>
<td>ERA1</td>
<td>ROP10</td>
<td>Partially promotes</td>
<td>Arabidopsis</td>
<td>[177]</td>
</tr>
<tr>
<td>ABH1</td>
<td>ABA → Ca&lt;sup&gt;2+&lt;/sup&gt;</td>
<td>Partially inhibits</td>
<td>Arabidopsis</td>
<td>[186]</td>
</tr>
<tr>
<td>ABH1</td>
<td>ABA → Closure</td>
<td>Partially inhibits</td>
<td>Arabidopsis</td>
<td>[186]</td>
</tr>
<tr>
<td>ABH1</td>
<td>AnionEM</td>
<td>Partially inhibits</td>
<td>Arabidopsis</td>
<td>[185]</td>
</tr>
<tr>
<td>CalM</td>
<td>ABA → KOUT</td>
<td>Partially promotes</td>
<td>Commelina communis</td>
<td>[358]</td>
</tr>
<tr>
<td>Ca&lt;sup&gt;2+&lt;/sup&gt;</td>
<td>ABA → OST1</td>
<td>No relation</td>
<td>Arabidopsis</td>
<td>[226]</td>
</tr>
<tr>
<td>Ca&lt;sup&gt;2+&lt;/sup&gt;</td>
<td>Depolarization → KAP</td>
<td>Partially inhibits</td>
<td>Arabidopsis</td>
<td>[361]</td>
</tr>
<tr>
<td>Ca&lt;sup&gt;2+&lt;/sup&gt;</td>
<td>H&lt;sup&gt;+&lt;/sup&gt;ATPase</td>
<td>Inhibits</td>
<td>Vicia faba</td>
<td>[363]</td>
</tr>
<tr>
<td>Ca&lt;sup&gt;2+&lt;/sup&gt;</td>
<td>PLC → InsP3</td>
<td>Promotes</td>
<td>Arabidopsis</td>
<td>[364]</td>
</tr>
<tr>
<td>Event</td>
<td>Agent</td>
<td>Effect</td>
<td>Plant</td>
<td>Reference</td>
</tr>
<tr>
<td>-------</td>
<td>-------</td>
<td>--------</td>
<td>-------</td>
<td>-----------</td>
</tr>
<tr>
<td>( \text{Ca}^{2+} )</td>
<td>AnionEM</td>
<td>Promotes</td>
<td>Arabidopsis</td>
<td>[237]</td>
</tr>
<tr>
<td>( \text{Ca}^{2+} )</td>
<td>InsP6 ( \rightarrow ) CIS</td>
<td>No relation</td>
<td>Vicia faba</td>
<td>[229]</td>
</tr>
<tr>
<td>( \text{Ca}^{2+} )</td>
<td>KOUT</td>
<td>no relation</td>
<td>S. tuberosum</td>
<td>[350]</td>
</tr>
<tr>
<td>( \text{Ca}^{2+} )</td>
<td>ABA ( \rightarrow ) Actin disruption</td>
<td>Promotes</td>
<td>Commelina communis</td>
<td>[232]</td>
</tr>
<tr>
<td>( \text{Ca}^{2+} )</td>
<td>NO ( \rightarrow ) AnionEM</td>
<td>Promotes</td>
<td>Vicia faba</td>
<td>[107]</td>
</tr>
<tr>
<td>( \text{Ca}^{2+} )</td>
<td>KEV</td>
<td>Promotes</td>
<td>Commelina communis</td>
<td>[358]</td>
</tr>
<tr>
<td>ABI1</td>
<td>ABA ( \rightarrow ) RAC1 (ROP6)</td>
<td>Partially</td>
<td>Arabidopsis</td>
<td>[176]</td>
</tr>
<tr>
<td>ABI1</td>
<td>ABA ( \rightarrow ) AnionEM</td>
<td>Inhibits</td>
<td>Arabidopsis</td>
<td>[237, 238]</td>
</tr>
<tr>
<td>ABI1</td>
<td>ABA ( \rightarrow ) ROS</td>
<td>Inhibits</td>
<td>Arabidopsis</td>
<td>[100, 237]</td>
</tr>
<tr>
<td>ABI1</td>
<td>ABA ( \rightarrow ) ( \text{Ca}^{2+} )</td>
<td>Inhibits</td>
<td>Arabidopsis</td>
<td>[237]</td>
</tr>
<tr>
<td>Depolarization</td>
<td>KOUT</td>
<td>Promotes</td>
<td>Arabidopsis</td>
<td>[84]</td>
</tr>
<tr>
<td>Hyperpolarization</td>
<td>Depolarization ( \rightarrow ) KAP</td>
<td>Promotes</td>
<td>Arabidopsis</td>
<td>[361]</td>
</tr>
<tr>
<td>Hyperpolarization</td>
<td>( \text{Ca}^{2+} ) dynamics</td>
<td>No relation</td>
<td>Vicia faba</td>
<td>[352]</td>
</tr>
<tr>
<td>KOUT</td>
<td>ABA ( \rightarrow ) Closure</td>
<td>Promotes</td>
<td>Arabidopsis</td>
<td>[84]</td>
</tr>
<tr>
<td>AnionEM</td>
<td>ABA ( \rightarrow ) Closure</td>
<td>Promotes</td>
<td>Arabidopsis</td>
<td>[88]</td>
</tr>
<tr>
<td>AnionEM</td>
<td>ABA ( \rightarrow ) Closure</td>
<td>Partially promotes</td>
<td>Commelina communis</td>
<td>[86]</td>
</tr>
<tr>
<td>AtP2C-HA</td>
<td>ABA ( \rightarrow ) Closure</td>
<td>Partially inhibits</td>
<td>Arabidopsis</td>
<td>[30]</td>
</tr>
<tr>
<td>AtPP2CA</td>
<td>ABA ( \rightarrow ) Closure</td>
<td>Partially inhibits</td>
<td>Arabidopsis</td>
<td>[153]</td>
</tr>
</tbody>
</table>
Appendix A. Text S1: Detailed Justification for Each Boolean Transfer Function

\[ \text{SphK} = \text{ABA} \]
\[ \text{S1P} = \text{SphK} \]
\[ \text{GPA1} = (\text{S1P or not GCR1}) \text{ and AGB1} \]
\[ \text{PLD} = \text{GPA1} \]
\[ \text{PA} = \text{PLD} \]

ABA-induced phosphatidic acid production is mediated by the S1P pathway and heterotrimeric G proteins [165]. ABA activates SphK, which produce S1P, and heterotrimeric G proteins are downstream of the S1P signal [165]. ABA induced S1P production peaks at 5 minutes after ABA stimulus, then decreases to a level above the resting level [165]. This dynamic behavior implies a negative feedback acting on S1P production. As nothing is known about the putative S1P feedback inhibitor, and there is no evidence that nodes downstream of S1P have transient behavior, we do not explicitly include negative feedback into the rule describing S1P activity. AGB1 is the sole Arabidopsis Gβ subunit [158]. GCR1 is a negative regulator of the ABA and S1P signal in guard cells and it can interact with GPA1 [355]. The \textit{gcr1} mutant shows stronger ABA response than the wild type at moderate ABA concentrations, but interaction with GCR1 does not block G protein signaling, thus we model this negative regulation by an “or not” function. In \textit{Arabidopsis}, the only canonical G protein α subunit, GPA1, can interact with PLDα1 and activates phospholipase D activity [354]. Phospholipase D catalyzes the production of phosphatidic acid (PA). Although there is no direct experimental evidence for S1P-induced PLD activity, the above observations taken together imply that S1P activates PLD through GPA1.
and activated PLD catalyzes PA production. We assume that the precursors of SphK, PLD, and the activity of AGB1 are not limiting factors for the signal transduction process.

\[ \text{pH}_c^* = \text{ABA} \]

ABA induces cytosolic pH increase (pH$_c$) \cite{72}. It has been shown that pH$_c$ increase starts immediately after ABA stimulus \cite{231}; cytosolic pH reaches a peak value at 15 minutes then decreases steadily, stabilizing at a level above the resting level \cite{72}. Since nothing is known about the putative pH$_c$ feedback inhibitor, and there is no evidence that nodes downstream of pH$_c$ have transient behavior, we do not explicitly include negative feedback into this rule.

\[ \text{OST1}^* = \text{ABA} \]
\[ \text{ROP2}^* = \text{PA} \]
\[ \text{Atrboh}^* = \text{pH}_c \text{ and OST1 and ROP2 and not ABI1} \]
\[ \text{ROS}^* = \text{Atrboh} \]

At least three pathways have been shown to regulate reactive oxygen species (ROS) production (catalyzed by Atrboh) in Arabidopsis. ABA failed to elicit ROS production in an ost1 mutant \cite{72, 226}. Our unpublished data and other indirect evidence also suggest that PA participates in the ABA-induced ROS production pathway \cite{102}. The third pathway is the cytosolic pH pathway; if cytosolic pH is clamped by butyrate, ABA-induced ROS production is also inhibited \cite{72}. ABA-induced ROS production is inhibited in the \textit{abi1-1} dominant mutant \cite{100}, we model the combination of activators (pH, OST1, ROP2) and inhibitor (ABI1) as an “and not” function.

\[ \text{H}^+ \text{ ATPase}^* = \text{not ROS and not pH}_c \text{ and not Ca}^{2+}_c \]
The H⁺ ATPase has been shown to be inhibited by reactive oxygen species [103], cytosolic calcium concentration increase [363] and cytosolic H⁺ concentration increase [365]. The “and” function reflects the assumption that any of the three negative regulators is sufficient to inhibit the H⁺ ATPase activity.

\[ \text{ABI1}^* = \text{pH}_c \text{ and not PA and not ROS} \]

ABI1 has been shown to be upregulated by cytosolic pH [146], and negatively regulated by ROS [104]. A recent paper shows that ABI1 is relocated from the cytosol to the plasma membrane by binding to phosphatidic acid [356], therefore PA negatively regulates ABI1 activity. In the equation, we use the “and” function to signify that the activation of ABI1 requires an increase in cytosolic pH and either negative regulator’s presence can efficiently reduce the strength of ABI1.

\[ \text{RCNI}^* = \text{ABA} \]
\[ \text{NIA12}^* = \text{RCN1} \]
\[ \text{NOS}^* = \text{Ca}^{2+}_c \]
\[ \text{NO}^* = \text{NIA12 and NOS} \]
\[ \text{GC}^* = \text{NO} \]
\[ \text{ADPRc}^* = \text{NO} \]
\[ \text{cADPR}^* = \text{ADPRc} \]
\[ \text{cGMP}^* = \text{GC} \]

These seven equations plus the CIS rule given below encapsulate the nitric oxide dependent pathway of induced intracellular Ca²⁺ release. ABA-induced NO production through the NOS pathway is calcium dependent [230], we hypothesize that nitric reductase (NIA12) activity is regulated by PP2A(RCN1) based on the well conserved NIA12 regulation mechanism [61]. Both cADPR and cGMP are downstream of the nitric oxide pathway in regulation of guard cell closure.
NIA and NOS can both mediate NO production. ADP ribose cyclase (ADPRc) catalyzes cADPR formation and guanyl cyclase (GC) catalyzes cGMP formation.

\[ \text{PLC}^+ = \text{ABA and Ca}^{2+}_c \]
\[ \text{InsP3}^+ = \text{PLC} \]
\[ \text{InsPK}^- = \text{ABA} \]
\[ \text{InsP6}^- = \text{InsPK} \]

ABA induce both InsP3 and InsP6 production in guard cells through the PLC [357] or InsPK pathway. PLC and InsPK represent, respectively, the known last steps in the catalysis of InsP3 and InsP6 production [229]. PLC-catalyzed InsP3 production requires Ca\(^{2+}\) [364].

\[ \text{CIS}^\dagger = (\text{cGMP and cADPR}) \text{ or (InsP3 and InsP6)} \]

In guard cells, multiple intracellular compartments are able to store Ca\(^{2+}\) and release it after stimulation. The “and” function between cGMP and cADPR reflects that these two signaling molecules are dependent and may cause Ca\(^{2+}\) release from the same compartment. Both InsP3 and InsP6 can induce Ca\(^{2+}\) release [357]. InsP6-caused Ca\(^{2+}\) release from intracellular stores is not Ca\(^{2+}\) dependent [229], and elicits a different type of Ca\(^{2+}\) increase compared to InsP3. However, the InsP3 and InsP6 pathways are biochemically linked [366], so we model their synergy as an “and” function. The “or” function between the cGMP/cADPR and Ins3/Ins6 pathways reflects the assumption that each of these pathways can efficiently induce Ca\(^{2+}\) increase inside the cell.

\[ \text{Ca}^{2+}_c \text{ ATPase}^+ = \text{Ca}^{2+}_c \]
\[ \text{Ca}^{2+}_c^- = (\text{CaIM or CIS}) \text{ and (not Ca}^{2+}_c \text{ ATPase)} \]
Cytosolic Ca\(^{2+}\) concentration can increase during ABA treatment due to Ca\(^{2+}\) influx from outside of the cell and/or Ca\(^{2+}\) release from certain cellular compartments. In reality, Ca\(^{2+}\) release from intracellular compartments can be mediated by different types of transporters localized on different intracellular membranes [95]; however, current experimental data do not provide enough evidence for a more detailed representation. We subsume the mechanisms that mediate Ca\(^{2+}\) efflux from the cytosolic compartment, including Ca\(^{2+}\) ATPases and Ca\(^{2+}\)/H\(^{+}\) antiporters, as the node Ca\(^{2+}\) ATPase in our model. This rule is the simplest way to incorporate Ca\(^{2+}\) oscillations which have been observed in guard cells [251, 274, 357].

\[
\text{AnionEM}^* = ((\text{Ca}^{2+}_c \text{ or } \text{pH}_c) \text{ and not ABI1}) \text{ or } (\text{Ca}^{2+}_c \text{ and } \text{pH}_c)
\]

Anion efflux through the membrane was shown to be upregulated by cytosolic Ca\(^{2+}\) and pH increase [25, 237]. Anion efflux current can be further divided into fast anion current (R type) and slow anion current (S type). Our rule only models the slow current that is not strongly regulated by membrane potential, because evidence only suggests the pH upregulation of slow anion channels [25]. In this rule, our hypothesis is that ABI1 is a negative regulator of anion efflux [237]. If ABI1 is present, both positive regulators (cytosolic Ca\(^{2+}\) and cytosolic pH) are required to activate the anion efflux; if ABI1 is not present, only one positive regulator is required to activate the anion efflux.

\[
\text{Depolar}^* = \text{KEV or AnionEM or (not H}^+\text{ ATPase) or (not KOUT) or Ca}^{2+}_c
\]

Generally speaking, the efflux of positive/negative ions out of the cytosol will hyperpolarize/depolarize the cell. We model K\(^{+}\) release from the vacuole to the cytosol and anion efflux and calcium influx across the plasma membrane as positive regulators of depolarization, and plasma membrane H\(^{+}\) ATPase and outwardly rectifying K\(^{+}\) efflux at the membrane as negative regulators of the depolarization process.
CaIM$^*$ = (ROS or not ERA1 or not ABH1) and (not Depolar)

Calcium influx across the plasma membrane requires both hyperpolarization and ROS production [367]. CaIM was hypersensitive to ABA in the era1 mutant, but the hyperpolarization-induced activation threshold of CaIM showed little change [181]. The abh1 mutant has a similar phenotype to era1, except only hypersensitivity in ABA induced Ca$^{2+}$ was reported. Here we hypothesize that ABH1 is regulating CaIM; we evaluated the possibility of ABH1 regulating CIS instead, and the results were similar.

KOUT$^*$ = (pH$_c$ or not ROS or not NO) and Depolar

Membrane depolarization will drive K$^+$ flow out of the cell, and the activity of outwardly rectifying K$^+$ currents is activated by cytosolic pH increase [225] and inhibited by ROS [362] and nitric oxide [109].

KAP$^*$ = (not pH$_c$ or not Ca$^{2+}$) and Depolar

The K$^+$ AP current is a transient K$^+$ current which was observed in Arabidopsis guard cells only [361]. The characteristic difference in the regulation of KAP is that, unlike KOUT, KAP is inhibited by both pH and cytosolic calcium. The activation of KAP also requires depolarization of the membrane potential.

KEV$^*$ = Ca$^{2+}$

Calcium-induced K$^+$ release through K$^+$-permeable channels in the tonoplast [368].

PEPC$^*$ = not ABA
Malate$^* = \text{PEPC and not ABA and not AnionEM}$

ABA inhibits PEPC activity in guard cells [233] and PEPC is an upstream enzyme of malate biosynthesis. ABA also regulates malate concentration by inducing malate breakdown [353]. Anion efflux negatively regulates intracellular malate concentration by releasing malate from the cytosol [369].

\text{RAC1}$^* = \text{not ABA and not ABI1}$

\text{Actin}$^* = \text{Ca}^{2+}$ or not \text{RAC1}$

Actin cytoskeleton reorganization has been shown to be a positive regulator of ABA induced stomatal closure, and dominant negative RAC1 expression causes actin cytoskeleton reorganization and stomatal closure [176]. ABI1 negatively regulates ABA inhibition of RAC1 [176].

\text{Closure}$^* = (\text{KOUT or KAP}) \text{ and AnionEM and Actin and not Malate}$

Stomatal closure requires both anion efflux from membrane [239] and potassium efflux via either slow outwardly rectifying $K^+$ channels [84] or fast transient (AP) $K^+$ channels [361]. Malate is a negative regulator of stomatal closure through its effect of osmotic regulation [353], and closure also requires the reorganization of the actin cytoskeleton [176]. The “or” function indicates that either $K^+$ current type is strong enough to drive closure; the “and” functions indicate that anion efflux, $K^+$ efflux, actin reorganization and malate breakdown are all required to drive efficient stomatal closure.
Appendix A. Text S2: Verification of the Inference Process and of the Resulting Network

Appendix A. Table S1 is based on human curation of published data, and consequently displays specific history-dependent features. For example, the earlier one component/process was published, the more frequently it will appear as a part of an entry in Appendix A. Table S1. The most frequently recorded result is “ABA induces closure” (35 times), all in the “process B” column; this is simply because “ABA induces closure” is a well-known fact and was tested in many experiments as a control to study other components in the process. Another example is that pH appears in 7 entries but Ca^{2+} appears in 29 entries, probably because the technique used in measuring intracellular Ca^{2+} is implemented in more laboratories than the technique used in measuring intracellular pH. To gauge the stability of Figure 3.2 (in the main text) with respect to the addition or removal of information from Appendix A. Table S1, we calculated the minimum number of entries involving a given node that need to be removed before we lose a path from ABA to closure mediated by that node. The table below indicates the total number of entries in Appendix A. Table S1 for network components predicted by our model to be important, and shows the minimum number of entries whose deletion is necessary to cause the loss of the ABA-closure path containing the component.

<table>
<thead>
<tr>
<th>Node</th>
<th>Minimum number of deleted entries</th>
<th>Total number of related entries</th>
<th>Ratio of previous two columns</th>
</tr>
</thead>
<tbody>
<tr>
<td>KOUT</td>
<td>5</td>
<td>10</td>
<td>0.5</td>
</tr>
<tr>
<td>Ca^{2+}</td>
<td>4</td>
<td>29</td>
<td>0.14</td>
</tr>
<tr>
<td>ROS</td>
<td>4</td>
<td>23</td>
<td>0.17</td>
</tr>
<tr>
<td>Depolar</td>
<td>4</td>
<td>13</td>
<td>0.30</td>
</tr>
<tr>
<td>PLD</td>
<td>4</td>
<td>6</td>
<td>0.66</td>
</tr>
<tr>
<td>pH</td>
<td>4</td>
<td>7</td>
<td>0.57</td>
</tr>
<tr>
<td>AnionEM</td>
<td>3</td>
<td>19</td>
<td>0.16</td>
</tr>
<tr>
<td>S1P</td>
<td>3</td>
<td>4</td>
<td>0.75</td>
</tr>
<tr>
<td>PA</td>
<td>3</td>
<td>7</td>
<td>0.42</td>
</tr>
<tr>
<td>ABI1</td>
<td>3</td>
<td>7</td>
<td>0.42</td>
</tr>
<tr>
<td>SphK</td>
<td>1</td>
<td>4</td>
<td>0.25</td>
</tr>
<tr>
<td>GPA1</td>
<td>1</td>
<td>4</td>
<td>0.25</td>
</tr>
<tr>
<td>Actin</td>
<td>1</td>
<td>3</td>
<td>0.33</td>
</tr>
</tbody>
</table>

Table S2. Perturbation analysis of Table S1. This table reveals a correlation between the number of interactions/relationships known about a node and the difficulty of removing a path from ABA to closure passing through that node by the elimination of table entries. The ratio between the minimum required information and the existing information does not seem biased, however.
Appendix A. Text S3: The Effect of Random Rewiring on the Network Dynamics

To gauge the uniqueness of our reconstructed network in reproducing ABA-induced closure, we test its dynamics’ resilience to random edge rewiring. As the interchange of a positive regulation with a negative regulation would change the sign of two paths in the model, which will definitely change the dynamics, we randomize the network in a more restricted way. We follow the rewiring methods used in [290] and we further preserve not only the degree but also the sign of the edges in the model. We gradually increase the degree of randomization from one pair of edges up to eight pairs of edges. After network randomization we test whether such a model can still show ABA-induced closure following the dynamic simulation methods described in the main text. We classify the effect of ABA-induced closure in the simulation after the probability of closure reaches a stationary value. If this value is 1 (100%), we name it “reaches unity”; this category corresponds to the normal, hyposensitive and hypersensitive responses we described in the main text. If the value is 0% we name it “reaches zero”, this category corresponding to the insensitive responses in the model. And if the stationary probability of closure is between 0 and 1, we name it “reduced”. Note that in this analysis we use only three of the five categories described in the main text because two of those categories (hypersensitivity and hyposensitivity) characterize the effect of disruption of nodes as compared to the normal response. We did 400 random rewirings for positive edge pairs, and 100 random rewirings for negative pairs, because there are twice as many positive edges in our model as negative edges. We summarize the results in the two bar graphs shown in Appendix A. Figure S1. As expected from the redundancy of the network, when only one pair of either positive edges or negative edges is randomized, in more than 70% of randomized models we still see ABA-induced closure (yellow bars). As the degree of randomization increases, the number of randomized networks that can produce ABA-induction...
of closure is reduced, whereas the number of reduced/insensitive networks increases. With randomized negative pairs, networks where ABA induces no closure (maroon bars) are more likely, in agreement with intuitive expectation.
Appendix B
Supplemental Information for Chapter 4

B.1. Publicly available datasets

Gene expression data from 73 human tissues and 61 mouse tissues with germa-condensed datasets [307] were downloaded from: http://wombat.gnf.org/index.html. For Arabidopsis gene expression data, raw CEL files from AtGenExpress (ATGE dataset) [308] were obtained from the ME00319 dataset at the TAIR ftp site (ftp://ftp.arabidopsis.org/home/tair/Microarrays/Datasets/ExpressionSet_ME00319/). The root dataset (ROOT dataset) was obtained from GSE5749 and GSE8934 datasets in the GEO database [309, 320]. We generated guard cell and whole leaf microarray datasets in our laboratory (GC dataset) [327]. A total of 237 raw CEL files from 80 tissues were analyzed for Arabidopsis. For rice, gene expression data for 42 tissue types [310] were obtained from the rice atlas project from: http://bioinformatics.med.yale.edu/riceatlas/overview.jspx.
B.2. Experimentally observed kurtosis and simulated kurtosis

We found that in all four organisms, human, mouse, Arabidopsis and rice, expression data showed heavy-tailed distribution of kurtosis (Figure 4.1 B). Kurtosis is calculated for each gene using the “moments” package (http://rss.acs.unt.edu/Rdoc/library/moments/html/00Index.html) with the following formula:

\[
K = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^4 \\
\left( \frac{1}{n} \sum_{j=1}^{n} (x_j - \bar{x})^2 \right)^2
\]

\(x_i\) is the gene expression in the \(i^{th}\) tissue type, and \(\bar{x}\) is the mean expression of the same gene. Gene expression levels for each gene were centered and scaled (Appendix B.3) before analysis. Such normalization does not affect the calculation of kurtosis and provides a way to define expression thresholds across multiple genes. In Figure 4.1B, the cumulative distribution of kurtosis of all genes in each of the four organisms is plotted as solid curves. In each simulation and for each gene, \(N\) normally distributed random numbers were generated, with \(N\) chosen to match the sample numbers in the microarray datasets in the four organisms respectively. For each gene, the median and median-absolute-deviation (MAD) of microarray data were used as parameters to generate normally distributed random numbers, and kurtosis of the simulated expressions was calculated. Median and MAD are used here because they are robust estimators of the location and spread of a distribution. Cumulative distribution of each simulated kurtosis was plotted as a dashed line in Figure 4.1B.
We ranked genes based on their expression kurtosis from the greatest to the least (Figure 4.1 A), and applied the Anscombe test [323], to identify genes that have excessive kurtosis (as measured by FDR adjusted p value (Q) < 0.005). In human, 15657 probesets out of 19854 probesets (78.8%), corresponding to 10665 genes out of a total of total 12837 (83.1%) genes as annotated by Ensemble on March 2009 (Bioconductor hgu133a.db version 2.2.12) exceed this Q threshold. In Arabidopsis, we find 14161 probesets out of 22810 probesets (62.1%), corresponding to 14528 genes out of a total of 23200 (62.6%) as annotated by TAIR downloaded on April 2009 (bioconductor ath1121501.db version 2.2.13), exceed this Q threshold. In mouse, we find 28900 probesets out of 36182 probesets (79.9%), corresponding to 13239 genes out of a total of 16018 (82.6%) as annotated by the original authors using Ensemble identifiers [307], exceed this Q threshold. In rice, we find 2503 probesets out of 6320 probesets (39.6%), corresponding to 2498 genes out of a total of 6275 (39.8%) as annotated by original authors [310], exceed this Q threshold.

We chose not to use other measurements of the shapes of distributions such as mean and variances or skewness, because they are less informative than kurtosis for selecting tissue preferential expression patterns. Genes with high average expression levels or high variation in expression levels do not necessarily have tissue preferential expression. We also noticed that most genes have positive skewness in expression (data not shown). Positive skewness is a weaker indicator for tissue preferential expression, because a distribution with positive skewness only means that the distribution has a central peak leaning towards lower expression levels.
B.3. Gene selection based on reduction of kurtosis

We define gene-tissue association using a threshold of the relative expression level $Z$, which is calculated by the following formula:

$$Z_{gi} = \frac{(E_{gi} - \text{MEDIAN}(E_g))/\text{MAD}(E_g)}{E_{gi}}$$

$E_{gi}$ is the expression level of gene $g$ in tissue $i$. $E_g$ is a vector of all gene expression levels in all tissue samples.

If a gene has a leptokurtic expression pattern, and high expression in some tissues is removed from the calculation of kurtosis for that gene, the expression distribution will have a reduced kurtosis. Therefore, we associate genes with tissues based on whether a given gene’s expression is higher than $Z$ in a particular tissue, then remove that gene expression in that tissue and calculate kurtosis again. If we use an arbitrarily large $Z$, moderately expressed genes will not be associated with tissues, even though their expression distribution can be still leptokurtic. On the other hand, using a small $Z$ can significantly decrease the kurtosis calculated from expression levels below $Z$, but gene-tissue associations identified by such small $Z$ thresholds will be less specific. To obtain a $Z$ threshold that balances the above two issues, we plotted the cumulative reduction of kurtosis (ROK) curve for any given kurtosis threshold using several different $Z$ threshold values. We then chose a threshold for $Z$ such that for 99% of genes, after removing the expression values above the threshold, expression kurtosis was reduced. In this way, we obtain a threshold that is dependent on all data available. We define FDR adjusted p value as $Q$ and by using this method, we found that for $Q < 0.005$, the corresponding $Z$ threshold is 3, and for $Q <
1e-09, the corresponding Z threshold is 10, for both human and Arabidopsis (Appendix B. Figure S1, A to D).
B.4. Leptokurtic distribution cannot be explained by a simple model of tissue specific expression

We centered gene expression by median and MAD to obtain a relative expression level (Z-score). To understand the basis of the kurtotic distributions, we identified genes with high kurtosis (high kurtosis genes; HKGs, Q<1e-9) for Arabidopsis and human, and examined histograms of gene expression for these genes (data not shown). We found that many genes have the following expression pattern: while the maximum expression level is observed in one single tissue type, the second maximum expression level is not as high as the maximum expression, but is still very high as compared to the expression levels in the remaining tissues.

Because for each gene, the maximum expression is only one number, we could not obtain an estimate of the variation of that number. Therefore, we quantify the distance of the 2nd highest expression level from the highest expression level by a range statistic (R_{12}). For each gene, we first rank the expression levels from high to low, with E₁ representing highest expression, and E₂ representing 2nd highest expression and Eₙ representing lowest expression, with n being equal to the number of tissue samples. Then R_{12} is defined as:

\[ R_{12} = \frac{E_1 - E_2}{E_1 - E_n} \]

\( R_{12} \) approaches 0 if the 2nd maximum expression level is close to the 1st expression level, and approaches 1 if the 2nd maximum expression level is close to the center of the rest of gene expression levels. To further characterize the expression patterns, we calculated \( R_{xy} \) values for all HKGs (Q<1e-9) using the following formula:
\[ R_{xy} = \frac{(E_x - E_y)}{(E_x - E_n)} \]

We restricted \( y = x - 1 \), such that only gaps between consecutive expression levels were calculated. We first calculated \( R_{xy} \) up to \( R_{9,10} \), and found that \( R_{xy} \) values after \( R_{45} \) are all close to zero. Therefore, for each gene, we only use \( R_{12} \) to \( R_{45} \) for further analysis.

For all HKGs that are associated with 4 or more tissues, we calculated vectors of \( R_{xy} = (R_{12}, R_{23}, R_{34}, R_{45}) \) for each gene, and carried out hierarchical clustering analysis of these genes based on \( R_{xy} \). We obtained 4 representative clusters for human HKGs and 5 representative clusters for Arabidopsis HKGs, and we plotted their mean \( R_{xy} \) profile and mean \( R_{xy} \pm \) standard deviation in Appendix B. \textbf{Figure S2 A and B}. As negative references, we also obtained \( R_{xy} \) for 1000 randomly selected genes with low kurtosis (\( Q > 0.005 \)), and plotted their mean profiles in Appendix B. \textbf{Figure S2 A and B} “low kurtosis” panel. To show an example of the expected \( R_{xy} \) patterns for genes with multiple hierarchies of expressions, we generated simulated expressions using the following random number generator:

\[ N = \{ N_1(\mu_1, \sigma_1), N_2(\mu_2, \sigma_2), N_3(\mu_3, \sigma_3), N_4(\mu_4, \sigma_4), N_5(\mu_5, \sigma_5) \ldots N_F(\mu_0, \sigma_0) \} \]

Each \( N_i \) is a random number generated using a Gaussian random number generator with parameters \( \mu \) and \( \sigma \).

\[ \mu_0 = 1, \sigma_0 = 1, \mu_1 = 5, \sigma_1 = 5, \mu_2 = 25, \sigma_2 = 10, \mu_3 = 125, \sigma_3 = 20, \mu_4 = 625, \sigma_4 = 40 \]
These parameters were chosen such that each simulated gene has four hierarchies of expression levels, and a higher hierarchy has a 5 fold increase in mean expression levels over the next level. $\sigma$s were chosen to reflect an increasing variation in expression level with increasing gene expression as observed in multiple organisms [370]. The parameters are not calculated from experimental data, because most genes do not show four levels of hierarchies. Instead, these parameters are used to illustrate the expected patterns of $R_{xy}$ if there are four hierarchies of expression levels. To match number of tissues in our datasets, we chose $T = 73$ for human and $T = 80$ for Arabidopsis. Because for all clusters in both human and Arabidopsis, mean profiles of $R_{12}$ to $R_{45}$ values show intermediate patterns between low kurtosis genes without hierarchical expression patterns and genes with 4 levels of hierarchy, we conclude that these HKGs exhibit a multi-level hierarchy of gene expression levels.
B.5. Tissue networks

In each gene centric tissue network (GCTN), nodes represent tissue types, with size of the node proportional to the number of leptokurtic genes associated with that tissue. Two nodes are connected with an edge if the two tissues share a significantly higher number of leptokurtic genes than expected by random as determined by a tissue similarity score (S). For any pair of tissues, a p value of Fisher’s exact test was calculated based on the number of genes associated with each tissue, number of genes associated with both tissues, and the sum of common and unique genes between the two tissues. FDR correction was used on all calculated p values to obtain a q value (FDR adjusted p value). The tissue similarity score (S) is defined as the negative log 10 of the q value. The maximum value of S is 320 due to the limitation of computing precision. The networks in Figures 4.2 and 4.3 were generated using $Q < 1 \times 10^{-9}$, $Z = 10$ and $S = 20$ to limit the number of edges that would appear in the whole figure for visual clarity. At $Q < 1 \times 10^{-9}$, the $Z$ and $S$ thresholds were determined by the ROK method and by inspecting graph theoretical properties in Figure 4.4, respectively. However, when we studied the graph theoretical properties for both human and Arabidopsis networks, we used both sets of thresholds.

We tested whether the commonly used co-expression method [325] generates networks for human and Arabidopsis that are similar to our GCTNs. We generated all the pair-wise tissue similarity scores using Pearson correlation coefficient from gene expression levels of all genes, and selected a threshold in the correlation matrix such that tissue pairs are connected if the correlation coefficient is above the threshold. We found that, for the human co-expression network, 0.07 is the threshold that would have to be used to obtain the same number of edges as our human GCTN, and 0.34 is the corresponding threshold for the Arabidopsis co-expression network. Neither of these values can be considered as high correlation. For these two co-
expression networks, we also observed that 66% of human GCTN edges and 75% of Arabidopsis GCTN edges are different from those in the correlation network. In co-expression networks, the expression of a group of selected genes is used to derive all pair-wise tissue similarities [308, 310]. By contrast, our approach will assign a unique set of leptokurtic genes that defines the similarity of each pair of tissues. This is the reason that GCTNs and co-expression networks have different topology and capture different properties of the expression data.
B.6. Text mining and comparison with manually annotated tissue expression patterns

The UniProt knowledgebase (UniProtKB) consists of two separate databases [324]: Swiss-prot and TrEMBL. The Swiss-prot database provides manually curated and reviewed information about tissue specificity from wet bench experiments in the “Tissue Specificity” section. Manual annotations were chosen for the comparison under the assumption that they are the most reliable, as they are based on targeted wetbench analyses. UniProtKB was downloaded on May 2009. This database is, to our knowledge, the most complete manually-curated database providing information about tissue specificity of human genes. Because the database is manually curated, the tissue specificities are described in sentences or short phrases, which cannot be readily matched to the annotations of microarray experiments. To validate our prediction of gene activity from microarray data, we constructed an automatic text mining system to extract the information about tissue specificity from the UniProtKB and then compared such information to our own.

To construct an automatic text mining system, we first created a mapping between the sample annotation in the microarray and the annotations used in the UniProtKB database. For most entries, mappings are straightforward. For example, “X721 B lymphoblasts” in the microarray annotation is mapped to lymphoblasts in the UniProtKB annotation. The ‘Adipocyte’ annotation in microarray data is mapped to ‘adipocyte’ or ‘adipose tissue’ in the UniProtKB annotation. Some annotations are so mapped due to the common shorthand for certain cell types. For example, CD4.T.cells and CD8.T.cells are separate samples in microarray experiments, but in the UniProtKB annotation, ‘CD4 and CD8-positive T- cell’, ‘T-cell’ and ‘T cells’ all refer to similar cell types measured by both microarray samples. Therefore, mappings between all these phrases were created so that each microarray tissue sample could map to one or more phrases in the
UniProtKB annotations, while each phrase in the UniProtKB annotations could map to one or more phrases in the microarray annotation. This mapping is the central element of our text mining system, and we call this list the 'known noun list'.

Automatic text mining is still not able to capture the semantics of long documents. However, automatic text mining on the UniProtKB has several advantages over text mining on publication abstracts or text from other secondary databases such as the OMIM database. First of all, the context of text (tissue specificity) is manually curated, thus it contains minimum noise from other biological information such as protein activity and protein-protein interactions. Second, in many texts in Pubmed abstracts or databases such as OMIM records, unrelated tissue types will also be mentioned, which could potentially decrease the specificity of text mining results. Third, the text of UniProtKB entries is relatively short; most of the text (93%) contains only 1 or 2 phrases to 1 or 2 sentences with less than 40 words. Because of these advantages, the results from text mining of UniProt database are a good benchmark resource for our analysis.

To map the UniProtKB annotation to microarray data, we downloaded probe to UniProt annotation of the hgu133a.db_2.2.11 annotation package from the bioconductor website (http://bioconductor.org/). In total, 18458 probesets map to UniProt IDs, and only those 9374 probesets that map to a single uniprot ID were used for the analysis. Among these 9374 probesets, 2342 probesets have Q<1e-9. For each description of a gene in UniProtKB, we calculated a similarity score, which is defined as the following:

\[
\text{Text Mining Similarity Score (TMSS)} = \left( \frac{M}{NM} \right) \times \left( \frac{M}{NU} \right)
\]

where:

- \(NM\) = Number of tissues selected by microarray
- \(NU\) = Number of tissues described in the UniProtKB annotation
- \(M\) = Number of matches between the two annotations
The TMSS is between one and zero. Any given gene will have a score of one if all identified nouns in the UniProtKB annotation exactly match the microarray annotation. This score will decrease for cases in which two annotations only partially match each other, and will be zero for a complete mismatch. To compare the effectiveness of kurtosis selection, we first assign each probeset (in the UniProt mapped subset) in the microarray to tissue types such that Z is > 10 in the selected tissues but not in other tissues. This corresponds to tissue specific genes without considering expression kurtosis.

For the text mining results, we compared our high kurtosis genes (HKG) (\(Q<1\times10^{-9}\)) method to two widely used methods. The first method uses ‘present call’ to determine gene-tissue associations. A gene will be associated with a tissue if that gene is called “present” in all microarray replicates in that tissue [307]. The second method uses an absolute expression threshold: a gene is associated with a tissue if the expression of that gene is above a given threshold in that tissue. We chose to use 200 as the threshold as suggested by the original publication [307] and 400 as an even more stringent threshold than the threshold of 270 suggested by a subsequent publication [315]. For all methods tested, we first assigned genes to tissues, and then we divided these genes into three groups: only associated with one tissue, associated with 2 to 10 tissues, and associated with more than 10 tissues. For each category, we compared the average TMSS of our method to other published methods using a Mann-Whitney test. The result is shown in Appendix B. Figure S3. We can see that with the threshold level we used to select HKGs, TMSS scores are significantly higher by our method than those obtained with the previously published methods.
B.7. Limitations of current data sources

Because available datasets contain cell types, tissues and organs, sometimes, a whole organ and a single cell type in that organ are both used in our analysis. This mixture of sample types and the resulting sample nesting effect could be one source of the observed multi-level expression patterns for leptokurtic genes and could contribute to some connections in the GCTNs. Nevertheless, we could confirm that many leptokurtic genes have multi-level expressions that result from completely non-overlapping tissues, e.g. liver and kidney. We also manually annotated potential nesting tissue pairs based on their annotations, and determined the number of connections that could arise due to tissue nesting. We found that, in the human GCTN which has 443 connections (Figure 4.2), only 76 (18%) of all connections could be due to tissue nesting effects, while 82% of connections are not due to tissue nesting. From these 76 connections, 37 connections are from densely connected modules such as connections between whole brain samples and samples from subsections of the whole brain (module 3), and connections between whole blood samples and blood cells (module 6). In the Arabidopsis GCTN (Figure 4.3) which has 476 connections, 73% of connections are not due to tissue nesting. Only 128 (27%) connections could be due to tissue nesting, with 115 of these arising from the densely connected root module (Figure 4.3 module 1). In addition to the 128 connections, 17 connections are from consecutive developmental stages, such as whole floral organs in different stages. According to the original annotation [308], these tissues are not nested. Removing these manually identified nesting connections does not change the overall structural properties of the GCTN of either species (data not shown). Connections in GCTNs derived from those leptokurtic genes associated with non-nested tissues are more biologically intriguing than those due to nested tissues. However, for a leptokurtic gene that is associated with two or more nested tissues, we cannot rule out the
possibility that this leptokurtic gene appears moderately active when the whole organ is assayed, but is actually activated at a higher level in one specific cell type in the same organ. The leptokurtic genes that originate from either tissue nesting or non-nesting effects are all important candidates for further experimental validation, but for different hypotheses.
B.8. Graph theoretical properties of human and Arabidopsis tissue networks

The maximum value of tissue similarity score \( S \) is 320 due to the limitation of computing precision. The tissue networks in Figures 4.2 and 4.3 in the main text were obtained using \( Q = 1e-09, Z = 10 \) and \( S = 20 \). These results are thus a snapshot at a specific combination of \( Q, Z \) and \( S \). However, the network and biological properties discussed in the main text are actually consistent over a wide range of threshold combinations. To demonstrate this fact, we plot, in Figure 4.4, four graph theoretical properties of human and Arabidopsis tissue networks: total number of connected components, total number of edges in the network, total number of edges in the giant connected component, and number of nodes in the giant connected component, at increasing \( S \), with \( Q = 1e-09, Z = 10 \). We found that both human and Arabidopsis networks have similar properties across a wide range of thresholds. Thus, Appendix B. Figure S4 shows the same properties at \( Q = 0.005, Z = 3 \). In both Figure 4.4 and Appendix B. Figure S4, panel A shows the number of connected components as a function of \( S \). Connected components are defined as distinct connected subgraphs with at least two nodes. This is different from the classical definition of connected components where single nodes without any connections are also counted as a component. We choose an alternative definition because we focus on the number of genes shared between tissues, so that disconnected nodes are not recounted. Panel A illustrates that over a wide range of \( S \) (from 1 to 20 for \( Q = 1e-9 \) and \( Z = 10 \) (Figure 4.4), from 1 to 50 for \( Q = 0.005 \) and \( Z = 3 \) (Appendix B. Figure S4), tissue networks are fully connected with one giant connected component. Panel B shows that the total number of edges in the graph decreases exponentially with increasing \( S \). Panel C shows the same pattern for number of edges in the giant connected component of the networks for each species. Panel B and panel C both suggest that pairs of tissues with common genes have a similar decreasing pattern in both human and
Arabidopsis. Panel D shows that the total number of nodes in giant connected components decreases with increasing \( S \). Note the sudden decreases in number of nodes around \( S = 30 \) for Figure 4.4 panel D and around \( S = 100 \) to 150 for Appendix B, Figure S4, panel D. In both giant connected components of human and Arabidopsis networks, number of nodes showed sudden drops as \( S \) increases, which suggest the breakdown of the tissue networks into two disconnected networks, each of similar sizes.
B.9. Module finding algorithms

We tested three different community finding algorithms, a fast greedy algorithm [371], a random walk based algorithm (http://arxiv.org/abs/physics/0512106) and a weighted label propagating algorithm [372], as implemented in the igraph package (http://igraph.sourceforge.net/index.html). These algorithms were selected because they can detect community in weighted graphs. The first two algorithms terminate based on the criterion of maximum modularity measurement, while the last algorithm is based on numerical simulation. We found that the label propagating algorithm works best on our dataset for two reasons. First, this algorithm assigns each node to modules based on module identity of its first neighbors. In our tissue network, nodes share common genes with their first neighbors but not necessarily with their second neighbors. Second, because this algorithm is not based on modularity of the whole network, smaller modules that have strong local connections (i.e. for the human network, module 5 with adrenal gland and adrenal cortex) can still be found by the algorithm while such modules will be undetectable by other algorithms. The ability to find small modules is important for our datasets because some specific tissue types are only sampled in a small number of experiments (e.g., two tissues for human adrenal gland module).
B.10. Number of genes with leptokurtic expression in human and Arabidopsis tissue networks

Median numbers of leptokurtic genes in the tissues of each module were plotted in Figure 4.5 A and B for human and Arabidopsis networks at different Q and Z thresholds. For both human and Arabidopsis, the number one threshold in Figure 4.5 A and B corresponds to $Q = 0.005$ and $Z = 3$; the number eight threshold corresponds to $Q = 1e-09$ and $Z = 10$. Other thresholds are linear interpolations of these two sets of Q and Z values. From Figure 4.5 A we can see that human module 6, blood cells, has a consistently higher median number of genes per module, followed by module 5 (testis), module 4 (muscular, respiratory and endocrine systems), and module 7 (liver and kidney). Module 1 (ganglions) consistently has the lowest median number of genes, while module 2 (adrenal gland and adrenal cortex) and module 3 (nervous system), are next to lowest in terms of the median numbers of genes. For the Arabidopsis network in Figure 4.5 B, Module 2 (seeds) has the highest median number of genes over most thresholds tested, followed by module 1 (roots), module 8 (floral organs and siliques) and module 7 (shoot apex). Module 4 (stems), module 3 (seedling green parts on MS plates), and module 6 (leaves) have smallest median numbers of genes.
B.11. More examples of high tissue similarity scores within modules in Arabidopsis

For Arabidopsis module 8, sepals from stage 12 (ATGE34) and stage 15 (ATGE41) flowers have $S = 162$, and siliques (fruits) with seeds of stage 4 (ATGE77) or stage 5 (ATGE78) have $S = 320$. However, the median of the four $S$ scores between ATGE34/41 and ATGE77/78 only equals 74.5. Another example is from Arabidopsis root cells (module 1). Xylem pole pericycle (J0121) and phloem pole pericycle (S17) are both specific types of pericycle cells, and root hair cells (COBL9) and Non-hair cells (GL2) are both specific types of epidermal cells. The $S$ score between J0121 and S17, and between COBL9 and GL2, equals 320 and 299 respectively. However, the median $S$ score between pericycle cells and epidermal cells is only 179.5.

We found that many enriched GO annotations agree well with known functions of the corresponding tissues. As an additional example to the seed example in the main text, we found that, in the shoot apex subnetwork, genes annotated as ‘response to jasmonic acid’ and ‘salicylic acid’ (GO:0009753 and GO:0009751) are enriched in the earlier stages of the shoot apex and gradually disappear in early flowers and late stage flowers. Positive regulators of flower development (GO:0009911) become enriched in early flower samples, and are reduced in late stage flower samples. In late stage flower samples, sexual reproduction function (GO:0019953) becomes enriched (Appendix B, Figure S6), consistent with the fact that reproduction occurs in the mature flower.
B.12. Gene ontology analysis

Gene ontology enrichment analysis was carried out using the eliminate count method in topGO software [330] and annotation package hgu133a.db_2.2.12 and ath1121501.db_2.2.13 in Bioconductor version 2.4 (http://bioconductor.org/packages/2.4/data/annotation/). The eliminate count method relies on the graph structure of GO annotations for the calculation of multi-comparison adjusted p values for GO enrichment analysis and we found this method makes more biological sense than traditional multiple test adjustment methods. In Table 4.2, **Annotated** = number of genes annotated by this GO term in the relevant genome; **Selected** = number of leptokurtic genes annotated by each GO term (this column is called “Significant” by the original topGO software, but is changed here to “Selected” for clarity); **Expected** = expected number of genes annotated by this GO term; **p value** is calculated using the eliminate count method with p value of 0.01 as an eliminating parameter; **Annotation** = a short description of each GO term. All enriched GO terms with p value < 0.01 were obtained. For GO terms with p < 0.01, one can sort these GO terms based on either p value (from smallest to largest) or based on the number of leptokurtic genes annotated by each GO term (“Selected column”), from greatest to the least. We found that if we sort GO terms based on p value, many GO terms that have smaller p values are more likely to be non-generic terms. However, we are interested in considering the relatively generic functions of HKGs, therefore, we ranked GO terms with p value < 0.01 from greatest to least based on the “Selected” column. This way of ranking GO categories guarantees that generic terms (annotated to many genes) are included such that our interpretation of GO enrichment captures generic biological functions. The top 10 GO annotations with the maximum number of “Selected” leptokurtic genes among all other significant GO annotations are shown in Table 4.2.
In this table, only one GO category (extracellular region GO:0005576) is enriched in both human and Arabidopsis HKGs (Q < 1e-9).

Other GO analyses in Appendix B. Figures S5, S6 and S7 were also carried out using the topGO method described above. However, when analyzing GO enrichment in specific cell types, such as guard cells, we have no preference for generic terms over other, less generic terms, because both generic terms such as “endomembrane system (GO:0012505)” and less generic terms, such as “response to chitin (GO:0010200)” provide hypotheses about cell specific functions. In the GO analyses of Appendix B. Figures S5, S6, and S7 the resulting eliminate-count p values were converted using negative log10 transformation such that a smaller p value becomes a larger number. Then, the transformed scores were used in hierarchical clustering of GO terms.

We further examined the HKGs associated with both fetal liver and its first-neighbor tissues using both GO annotation and the OMIM database. We carried out GO enrichment analysis for four groups of genes in the network: genes that are associated with adult liver and its unique first neighbors (group 1), genes that are associated with fetal liver and its unique first neighbors (group 2), genes that are associated with both adult and fetal liver and are common first neighbors (group 3), and genes associated with both adult and fetal liver (group 4) (Figure 4.6C and Appendix B. Figure S7). First of all, genes common to both adult liver and fetal liver have the largest number of enriched GO categories, and these are mostly related to the common metabolic functions of both fetal and adult liver. We also noticed that genes associated with adult liver and its first tissue neighbors are specifically enriched in genes related to gluconeogenesis (GO:0006094), which is well known to be a process specifically carried out in adult liver but not fetal liver and also in adult kidney cortex [373]. Genes associated with fetal liver and their unique first neighbors are especially enriched in the process of erythrocyte maturation (GO:0043249), which corresponds well with haematopoesis, which is specific to fetal but not adult liver [326]. We further examined genes with OMIM annotations in each group, and we found that genes such as SUOX (OMIM:606887) and ASL (OMIM:608310) appear in group 1. Both genes are related to disorders after birth, caused by mutations in metabolic genes that are expressed in adult liver and are not highly expressed in fetal liver [374, 375]. We also found EPOR (OMIM: 133100), a gene associated with erythrocytosis, in group 2.
B.14. Experimental procedures for guard cell microarrays and phenotypic assays.

B.14.1. Guard cell and leaf microarray data

Generation of guard cell and leaf microarray data by our laboratory is described in [327]. Array data have been submitted in a MIAME compliant fashion in the confidential section of the Gene Expression Omnibus (GEO) database. The data can be accessed by the GEO reviewer link http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?token=zxovhsoemnwogrg&acc=GSE19

520
B.14.2. Validation of guard-cell leptokurtic genes using quantitative real time reverse-transcriptase PCR (Q-PCR).

Total RNA was isolated from tissue samples using the TriZol reagent (Invitrogen, CA), treated with RNase free DNase I, and purified using the RNAeasy kit (Qiagen, CA). For quality control, each RNA sample was analyzed by Bio-analyzer (Agilent Technologies). Purified RNA samples were used for cDNA synthesis using the Superscript III first strand synthesis kit (Invitrogen, CA). cDNA was diluted at a concentration of 1:100, aliquoted, and kept at 4°C throughout each experiment to avoid discrepancy in the data because of freeze-thaw cycles. Q-PCR was performed using pre-mix containing SYBR-Green intercalating dye (BioRad). Actin gene was used as an internal control (Charrier et al., 2002). The position of the oligonucleotides used for real-time PCR was chosen so that the size of all PCR products was between 100 and 150 bp. The suitability of the oligonucleotide sequences in term of efficiency of annealing was evaluated in advance using the Primer 3 program. Q-PCR experiments were repeated thrice independently, and the data were averaged. The data obtained were analyzed with IQ5 software (Bio-Rad). Primers for each gene are given in Appendix B. Table S5.
B.14.3 Verification of T-DNA Insertional Mutants.

The T-DNA Express database (http://signal.salk.edu/cgi-bin/tdnaexpress; Alonso JM et al., 2003) was searched for the availability of T-DNA insertion lines in Arabidopsis genes showing leptokurtotic expression in guard cells. Four genes, At1g50400, At5g60410 (previously characterized and named siz1, Miura et al., 2005), At2g21080 and At1g11100 were selected for further characterization as they had at least two independent and homozygous insertion lines available at the time of the analysis. T-DNA left border LBb1 and gene specific primers were used to confirm the reported insertion sites. Gene specific primers flanking each of the insertion sites were used to confirm the homozygosity of the insertions by genomic DNA PCR. DNA isolated from wild-type Col plants was used as controls in PCR reactions. The sequences of the primers used for PCRs are listed in Appendix B, Table S5.
B.14.4 Stomatal Assays

For stomatal aperture measurements, plants were grown under identical conditions as for guard cell isolation for microarray analysis [327]. For each sample two fully expanded leaves from 5 week-old plants were harvested just before the beginning of the light period. Excised leaves were placed in a six well Petri dish, abaxial side down in 5 mL of buffer solution. The composition of the solution for assays of ABA inhibition of stomatal opening was 10 mM KCl, 7.5 mM IDA, and 10 mM MES (pH 6.15) and the composition of the solution for assays of ABA promotion of stomatal closure was 20 mM KCl, 5 mM MES, and 1 mM CaCl₂ (pH 6.15). Excised leaves from wild type Col and the homozygous mutant plants were kept in darkness for 2 h and then transferred to light (450-µmol·m⁻²·s⁻¹) for 2.5 h in the presence of ABA (50 µM) to study inhibition of opening. For promotion of closure experiments, excised leaves were first kept in light for 2 h followed by addition of ABA (50 µM) and further incubation in light for 2.5 h. Cover slips (two per well) were added on top of the leaves to keep them evenly submerged in the buffer solutions.

The epidermes of the leaves were peeled after the end of incubation periods and mounted on a glass slide. The peels were photographed using a digital camera mounted to a Nikon Diaphot 300 microscope. Ten images were recorded per leaf (20 per sample). Apertures were measured using Image J software (rsbweb.nih.gov/ij), with a micrometer image (photographed with each sample) used as a scale. At least 100 apertures were measured for each sample. Three double blind biological replicates of the experiment were performed, and data were averaged.

For the measurement of stomatal lengths, the measurements were performed along the length of the aperture between the two guard cells. At least 100 aperture lengths were measured from non-ABA treated epidermal peels from three replicates and data were averaged.
B.15. Combining gene expression datasets using a linear model

Expression data from our arrays (Appendix B.10) and those of AtGenExpress (Schmid et al., 2005) and root data from two previous publications [309, 320] were normalized and summarized using the RMA algorithm [376]. Because datasets from Arabidopsis were obtained from multiple studies, to allow comparison of these datasets, we applied a nested linear model method to renormalize datasets across studies. We formulate the following gene-wise linear model based on the properties of the tissue samples used in different studies:

\[
Y_{ijk} = \alpha_i + \beta_{i(j)} + \epsilon_{i(jk)}
\]

In this model, \(\alpha\) represents study effect, \(\beta\) represents tissue effects and \(\epsilon\) represents error term. \(i = \text{ATGE or ROOT or GC}, j = \text{tissue samples}, k = \text{replicates}\).

For both ATGE and GC datasets, three replicates for each tissue were available and were used in our analysis. For the ROOT dataset, 2 to 4 replicates were used depending on availability. In both ATGE and GC datasets, mature rosette leaves were sampled. Guard cells were sampled only by our group, while some other tissues, including flowers and seeds and whole roots, were only represented in the ATGE dataset. We assume that any difference in gene expression obtained for a gene in leaves in the ATGE sample vs. our leaf samples reflects only the effect of different experimental conditions in different studies. This assumption is similar to the classical nested linear model. Specifically, we assume that the sum of \(\beta_{i(j)}\) for each \(i\) is zero, for similar tissue types. Based on this assumption, we can obtain the numerical estimation of study effects \(\alpha_i\) using only a subset of the data. For comparing ATGE and GC datasets, 25 samples \(j = \text{ATGE}_5, \text{ATGE}_7, \text{ATGE}_10, \text{ATGE}_12, \text{ATGE}_13, \text{ATGE}_14, \text{ATGE}_15, \text{ATGE}_16, \text{ATGE}_17, \text{ATGE}_19, \text{ATGE}_20, \text{ATGE}_21, \text{ATGE}_22, \text{ATGE}_23, \text{ATGE}_24, \text{ATGE}_25, \text{ATGE}_26,\)
ATGE_87, ATGE_89, ATGE_90, ATGE_91, ATGE_96, ATGE_97, ATGE_100, ATGE_101) from the ATGE dataset were used. These samples include rosette leaves and seedling green parts. For the GC dataset, data from mature rosette leaves were used. The above 25 samples from the ATGE dataset and the one sample from the GC dataset constitute the ATGE-GC sub-dataset. For ATGE and ROOT datasets, only whole root or root sections were used in the model so that seven samples (j = ATGE_3, ATGE_9, ATGE_93, ATGE_94, ATGE_95, ATGE_98, ATGE_99) for the ATGE dataset, and three samples (j = Stage I, Stage II, and Stage III) for the ROOT dataset were designated the ATGE-ROOT sub-dataset. For both subsets of data, least square estimations were obtained for $\alpha_i (i = \text{ATGE, ROOT, GC})$ for each gene. The differences between studies were calculated as the differences between corresponding $\alpha_i$. Expression levels in GC and ROOT datasets were adjusted toward the ATGE dataset by subtracting the differences between study effects:

$$Y_{ijk}^{adj} = \alpha_i - (\alpha_{i}^{\text{sub}} - \alpha_{i}^{\text{sub}'}) + \beta_{j(i)} + \epsilon_{i(j,k)}$$

$\alpha_{i}^{\text{sub}}$ stands for the GC dataset in the ATGE-GC sub-dataset, and for the ROOT dataset in the ATGE-ROOT sub-dataset. $\alpha_{i}^{\text{sub}'}$ stands for the ATGE dataset for both the ATGE-GC comparison and the ATGE-ROOT comparison, because the ATGE dataset contains both whole root data that are comparable to the ROOT dataset, and whole leaf data that are comparable to the GC dataset. The above described adjustment method is similar to published methods for adjusting batch effects [377], but incorporates the matched tissue samples such that samples measured only in one study but not in other studies can be compared.
B.16. Enrichment of human disease genes in HKG genes and comparison to tissue specific genes

The Online Mendelian Inheritance in Man (OMIM) morbidmap was downloaded from (http://www.ncbi.nlm.nih.gov/Omim/) on April, 2009. We specifically focused on genes that have OMIM category (3) annotation in the morbid map, because, according to the OMIM help file, the category (3) type of OMIM records have both the wild type gene mapped to a location and a known mutation in that gene that is associated with a disorder. Probesets and OMIM ID mapping were downloaded from the BioMart database (http://www.bioma-r.org/). This mapping was further filtered to keep only the 2076 probesets for which each probeset maps to one unique OMIM ID with category (3) in morbid map. Then we used the Fisher exact test to test whether the probesets that passed our kurtosis test threshold (Q) are enriched in probesets (genes) from the OMIM (3) datasets. For Q = 1e-9 (the peak of the line in Figure 4.8), we found that 4149 probesets passed our kurtosis test, and 694 of these probesets map to OMIM (3) genes, representing a significant enrichment of diseases genes (p value < 4.2e-45) in the HKGs.
B.17. Evolutionarily conserved genes between human and Arabidopsis.

Evolutionarily conserved genes were downloaded from the Inparanoid database [329]. The number of human HKGs (Q<1e-9) with Arabidopsis orthologous genes (these could be either HKGs or non-HKGs) according to the Inparanoid database was counted, and the statistical significance of enrichment was calculated using Fisher’s exact test. The same analysis was carried out for Arabidopsis HKGs. For those HKGs that are conserved, we counted how many of their orthologs are also identified as HKGs in the counterpart organism. The statistical significance was estimated by first finding the number of HKGs in one organism, and then randomly selecting the same number of genes from all genes measured in that organism. Then, the number of orthologs that are HKGs in the counterpart organism was counted. Such random selection was carried out 1000 times, and p value is calculated as the probability of randomly counted genes to be larger than the observed number.
Appendix B. Figure S1. Panels A to D: Reduction of kurtosis plots. We define Q as FDR adjusted p value from kurtosis test (Anscombe test), and we define relative expression levels as Z. The four panels in this figure show cumulative reduction of kurtosis across all leptokurtic genes with different thresholds of Q and Z. A. Human dataset, Q = 0.005, Z = 1,2,3,5,10 B. Arabidopsis dataset, Q = 0.005, Z = 1,2,3,5,10. C. Human dataset, Q = 1e-9, Z = 2,5,10,20,30. D. Arabidopsis dataset, Q = 1e-9, Z = 2,5,10,20,30.
Appendix B. Figure S2. Panel A and B: Clustering profiles of R12 to R45 of human and Arabidopsis HKGs. Only genes that have Q<1e-9 and are associated with four or more tissues were used for calculation of R12 to R45 (Rxy) values. Four clusters (panel A, group 1 to group 4) and five clusters (panel B, group 1 to group 5) were found for human and Arabidopsis respectively. Mean cluster profiles were plotted as solid lines, and mean cluster profiles +/- standard deviations were plotted as dashed lines. Profiles from human and Arabidopsis genes with “low kurtosis” (Q > 0.005, see Appendix B.4) and simulated genes with “four levels” of expression were also plotted.
Appendix B. Figure S3. Tissue associations of leptokurtic genes derived by our method perform better than gene-tissue associations from other published methods. Text mining similarity score (TMSS) is plotted for four different methods to define gene-tissue associations. **HKGs**: High kurtosis genes, genes with high kurtosis as defined by $Q<1e-09$ and $Z = 10$. **P**: gene-tissue associations defined by “Present/Absent Calls”. **200**: gene tissue associations defined by expression threshold of 200. **400**: gene tissue associations defined by expression threshold of 400. * $p$ value $< 0.01$, ** $p$ value $< 0.05$. See Supporting Information F for details.

Appendix B. Figure S4. Graph theoretical properties of tissue networks with $Q = 0.005$ and $Z = 3$.

Open circles: human tissue network. Open triangles: Arabidopsis tissue network. Four graph theoretical properties are plotted with different thresholds of: **A**, Number of connected components. **B**, Total number of edges in the network. **C**, Number of edges in the giant connected component. **D**, Total number of nodes in the giant connected component. For all four panels, x axis represents tissue similarity score ($S$).
Appendix B. Figure S5. GO annotation analysis for genes appearing in seed subnetwork. Stage 3, Stage 4 and Stage 5 are siliques that contain immature seeds. Stage 6, Stage 7, Stage 8, Stage 9 and Stage 10 are seeds without siliques. In the silique to seed subnetwork, pectinesterase (GO:0030599) and inhibitor of pectinesterase activity (GO: 0046910) are enriched in early silique development, while seed development (GO:0048316), negative regulation of seed germination (GO:0010187) and embryonic development ending in seed dormancy (GO:0009793) become enriched in later stages of seed development. High enrichment scores (negative log p value) are shown in red, and low to no enrichment are shown in yellow to white. Color bar shows the enrichment score (negative p value) in the heatmap. Enrichment scores >8 are rounded to 8 for clarity.
Appendix B. Figure S6. GO annotation analysis for genes appearing in shoot apex subnetwork. Vegetative Young Leaves, Vegetative, Before Bolting and Inflorescence are four samples from consecutive developmental stages of shoot apex, and Stage 9, Stage 10/11 and Stage 12 are flowers at consecutive developmental stages after Inflorescence. In the shoot apex subnetwork (denoted “Apex” in the original publication), genes annotated as ‘response to jasmonic acid’ and ‘salicylic acid’ (GO:0009753 and GO:0009751) are enriched in the earlier stages of the shoot apex and gradually disappear in early flowers, and late stage flowers. Positive regulators of flower development (GO:0009911) become enriched in early flower samples, and are reduced in late stage flower samples. Finally, sexual reproduction function (GO:0019953) becomes enriched in later stages of flower development, consistent with the fact that reproduction occurs in the mature flower.
Appendix B. Figure S7. GO analysis for human liver and fetal liver subnetwork. Adult liver and fetal liver common genes (Group 4) have the largest number of enriched GO categories, and these are mostly related to the metabolic functions of both fetal and adult liver. We also noticed that genes associated with adult liver and its first tissue neighbors (Group 1) are specifically enriched in genes related to gluconeogenesis (GO:0006094), which is well known to be a process specifically carried out in adult liver and also in adult kidney cortex [373]. Genes associated with fetal liver and their unique first neighbors (Group 2) are especially enriched in the process of erythrocyte maturation (GO:0043249), which corresponds well with fetal liver specific function [326].
Appendix B. Figure S8. Panel A: Tissue specific expression analysis of selected GC leptokurtic genes by Q-PCR. GC, Guard cells; LF, leaves; Rt, roots; St, stems; Sl, siliques; Fl, flowers; and Ct, cotyledons. Expression of each gene is plotted in comparison to its expression in guard cell cDNA which was set to 100%. Actin 2/8 gene was used as an internal normalization control and the PhyB gene (At5g36720; K = 4.49) was used as a negative control. The experiment was repeated thrice and data were averaged. Error bars represent S.E. At1g64010, putative serine protease inhibitor; At4g28460, unknown protein; At4g11330, MPK5; At4g16820, similar to DAD1; At5g35320, hypothetical protein; At2g21080, similar to extracellular ligand gated ion channel protein; At3g51760, hypothetical protein; At2g22320, unknown protein; At1g50400, similar to porin family protein; At1g11100, SNF2 domain containing protein; At5g60410, SIZ1; At2g19180, unknown protein; At2g27200, similar to GTP-binding protein; At1g20880, RRM containing protein. Kurtosis values for each gene are calculated from combined Arabidopsis microarray datasets using the kurtosis function in “moments” package from programming language R. Panel B. Schematic of the local genomic structure and T-DNA insertion sites of mutant Arabidopsis lines used for phenotypic analyses. The genomic structure of each gene is drawn to scale and the lengths of all gene model are normalized. Open bars indicate 5′ and 3′ untranslated regions, solid black bars indicate exons, and black lines indicate introns. T- DNA insertions are not drawn to scale. Panel C. RT-PCR showing absence of full length transcripts from the knockout mutant lines. Insertion sites and absence of full-length transcripts for these genes were confirmed by genomic PCR and RT-PCR, respectively. Gene-specific primers flanking the insertion sites were used in RT-PCR to confirm absence of full-length transcripts in the T-DNA insertional lines. Wild type Columbia was used as a positive control for each primer pair. Panel D. Percentage of different developmental stages of stomatal complexes in siz1 mutants versus wild type Columbia. Stomata were categorized into mature (fully developed), immature (ratio of aperture length to cell length less than 0.4), meristemoid (no defined aperture) and guard mother cell (non-divided cell). Two hundred stomatal complexes were scored for each sample, from 8 different leaves. Panel E. Stomatal phenotypes of the second independent T-DNA knockout alleles of the four mutants depicted in main text Figure 7. Images are from ABA-treated leaves. Superimposed numbers represent average cell length (n = 100 ± S.E.) Scale bar = 10 µm.
Appendix B. Figure S9. Root phenotypes of T-DNA insertional mutants of *At1g11100*.

Seeds of wild type and mutants plants (*At1g11100*) grown under identical conditions were used for assays of ABA-inhibition of root elongation. Open bars represent control (no ABA) and closed bars represent ABA (1 µM) treated roots at 8 days after germination. The experiment was repeated twice and data were averaged, ± S.E. For each experiment n=32 plants were measured per genotype per treatment. Numbers at the top represent percentage change in root length in the presence of ABA. Both alleles show significant reduction of root length as compared with wild-type Columbia (* P<0.01) in response to ABA.
### Appendix B. Table S1. Tissue annotations for Figure 4.2

Labels: node labeling in **Figure 4.2**. Color coding: tissues are colored coded according to their categories. File names: data label exactly as in the original dataset. **M**: module number in **Figure 4.2**. **N**: number of genes that are associated with each tissue.

<table>
<thead>
<tr>
<th>Labels</th>
<th>Color coding</th>
<th>File names</th>
<th>M</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Atrioventricular Node</td>
<td>Nervous</td>
<td>atrioventricularnode</td>
<td>1</td>
<td>31</td>
</tr>
<tr>
<td>Ciliary Ganglion</td>
<td>Nervous</td>
<td>Ciliary.ganglion</td>
<td>1</td>
<td>43</td>
</tr>
<tr>
<td>Dorsal Root Ganglia</td>
<td>Nervous</td>
<td>dorsal.root.ganglia</td>
<td>1</td>
<td>37</td>
</tr>
<tr>
<td>Trigeminal Ganglion</td>
<td>Nervous</td>
<td>Trigeminal.ganglion</td>
<td>1</td>
<td>35</td>
</tr>
<tr>
<td>Adrenal Cortex</td>
<td>Endocrine</td>
<td>AdrenalCortex</td>
<td>2</td>
<td>104</td>
</tr>
<tr>
<td>Adrenal Gland</td>
<td>Endocrine</td>
<td>adrenal gland</td>
<td>2</td>
<td>97</td>
</tr>
<tr>
<td>Amygdala</td>
<td>Nervous</td>
<td>Amygdala</td>
<td>3</td>
<td>201</td>
</tr>
<tr>
<td>Caudate Nucleus</td>
<td>Nervous</td>
<td>caudate.nucleus</td>
<td>3</td>
<td>78</td>
</tr>
<tr>
<td>Cerebellum</td>
<td>Nervous</td>
<td>Cerebellum</td>
<td>3</td>
<td>80</td>
</tr>
<tr>
<td>Cerebellum Peduncles</td>
<td>Nervous</td>
<td>Cerebellum.Peduncles</td>
<td>3</td>
<td>76</td>
</tr>
<tr>
<td>Cingulate Cortex</td>
<td>Nervous</td>
<td>CingulateCortex</td>
<td>3</td>
<td>73</td>
</tr>
<tr>
<td>Fetal Brain</td>
<td>Nervous</td>
<td>fetalbrain</td>
<td>3</td>
<td>222</td>
</tr>
<tr>
<td>Globus pallidus</td>
<td>Nervous</td>
<td>globus.pallidus</td>
<td>3</td>
<td>37</td>
</tr>
<tr>
<td>Hypothalamus</td>
<td>Nervous</td>
<td>Hypothalamus</td>
<td>3</td>
<td>146</td>
</tr>
<tr>
<td>Medulla Oblongata</td>
<td>Nervous</td>
<td>MedullaOblongata</td>
<td>3</td>
<td>83</td>
</tr>
<tr>
<td>Occipital Lobe</td>
<td>Nervous</td>
<td>OccipitalLobe</td>
<td>3</td>
<td>95</td>
</tr>
<tr>
<td>Parietal Lobe</td>
<td>Nervous</td>
<td>ParietalLobe</td>
<td>3</td>
<td>48</td>
</tr>
<tr>
<td>Pituitary</td>
<td>Nervous</td>
<td>Pituitary</td>
<td>3</td>
<td>112</td>
</tr>
<tr>
<td>Pons</td>
<td>Nervous</td>
<td>Pons</td>
<td>3</td>
<td>39</td>
</tr>
<tr>
<td>Prefrontal Cortex</td>
<td>Nervous</td>
<td>PrefrontalCortex</td>
<td>3</td>
<td>233</td>
</tr>
<tr>
<td>Spinal Cord</td>
<td>Nervous</td>
<td>Spinalcord</td>
<td>3</td>
<td>96</td>
</tr>
<tr>
<td>Subthalamic Nucleus</td>
<td>Nervous</td>
<td>subthalamicnucleus</td>
<td>3</td>
<td>57</td>
</tr>
<tr>
<td>Temporal Lobe</td>
<td>Nervous</td>
<td>TemporalLobe</td>
<td>3</td>
<td>47</td>
</tr>
<tr>
<td>Thalamus</td>
<td>Nervous</td>
<td>Thalamus</td>
<td>3</td>
<td>83</td>
</tr>
<tr>
<td>Whole Brain</td>
<td>Nervous</td>
<td>WholeBrain</td>
<td>3</td>
<td>160</td>
</tr>
<tr>
<td>Olfactory Bulb</td>
<td>Nervous</td>
<td>OlfactoryBulb</td>
<td>4</td>
<td>124</td>
</tr>
<tr>
<td>Bronchial Epithelial Cells</td>
<td>Respiratory</td>
<td>bronchial.epithelialcells</td>
<td>4</td>
<td>391</td>
</tr>
<tr>
<td>Fetal Lung</td>
<td>Respiratory</td>
<td>fetal.lung</td>
<td>4</td>
<td>342</td>
</tr>
<tr>
<td>Lung</td>
<td>Respiratory</td>
<td>Lung</td>
<td>4</td>
<td>368</td>
</tr>
<tr>
<td>Trachea</td>
<td>Respiratory</td>
<td>Trachea</td>
<td>4</td>
<td>201</td>
</tr>
<tr>
<td>Cardiac Myocytes</td>
<td>Muscular</td>
<td>CardiacMyocytes</td>
<td>4</td>
<td>280</td>
</tr>
<tr>
<td>Heart</td>
<td>Muscular</td>
<td>Heart</td>
<td>4</td>
<td>220</td>
</tr>
<tr>
<td>Skeletal Muscle</td>
<td>Muscular</td>
<td>SkeletalMuscle</td>
<td>4</td>
<td>124</td>
</tr>
<tr>
<td>Smooth Muscle</td>
<td>Muscular</td>
<td>SmoothMuscle</td>
<td>4</td>
<td>364</td>
</tr>
<tr>
<td>Tissue</td>
<td>Type</td>
<td>Code</td>
<td>Page</td>
<td>Value</td>
</tr>
<tr>
<td>--------------------------------</td>
<td>------------</td>
<td>--------------</td>
<td>------</td>
<td>-------</td>
</tr>
<tr>
<td>Tongue</td>
<td>Muscular</td>
<td>TONGUE</td>
<td>4</td>
<td>179</td>
</tr>
<tr>
<td>Fetal Thyroid</td>
<td>Endocrine</td>
<td>fetalThyroid</td>
<td>4</td>
<td>147</td>
</tr>
<tr>
<td>Pancreas</td>
<td>Endocrine</td>
<td>Pancreas</td>
<td>4</td>
<td>150</td>
</tr>
<tr>
<td>Pancreatic Islets</td>
<td>Endocrine</td>
<td>PancreaticIslets</td>
<td>4</td>
<td>257</td>
</tr>
<tr>
<td>Prostate</td>
<td>Endocrine</td>
<td>Prostate</td>
<td>4</td>
<td>321</td>
</tr>
<tr>
<td>Thyroid</td>
<td>Endocrine</td>
<td>Thryoid</td>
<td>4</td>
<td>387</td>
</tr>
<tr>
<td>Appendix</td>
<td>Digestive</td>
<td>Appendix</td>
<td>4</td>
<td>87</td>
</tr>
<tr>
<td>Salivary Gland</td>
<td>Digestive</td>
<td>salivarygland</td>
<td>4</td>
<td>103</td>
</tr>
<tr>
<td>Adipocyte</td>
<td>Others</td>
<td>ADIPOCYTE</td>
<td>4</td>
<td>284</td>
</tr>
<tr>
<td>Skin</td>
<td>Others</td>
<td>skin</td>
<td>4</td>
<td>72</td>
</tr>
<tr>
<td>Ovary</td>
<td>Reproductive</td>
<td>Ovary</td>
<td>4</td>
<td>58</td>
</tr>
<tr>
<td>Placenta</td>
<td>Reproductive</td>
<td>PLACENTA</td>
<td>4</td>
<td>510</td>
</tr>
<tr>
<td>Uterus</td>
<td>Reproductive</td>
<td>Uterus</td>
<td>4</td>
<td>263</td>
</tr>
<tr>
<td>Uterus Corpus</td>
<td>Reproductive</td>
<td>UterusCorpus</td>
<td>4</td>
<td>109</td>
</tr>
<tr>
<td>Testis</td>
<td>Reproductive</td>
<td>Testis</td>
<td>5</td>
<td>183</td>
</tr>
<tr>
<td>Testis Germ Cell</td>
<td>Reproductive</td>
<td>TestisGermCell</td>
<td>5</td>
<td>232</td>
</tr>
<tr>
<td>Testis Interstitial</td>
<td>Reproductive</td>
<td>TestisInterstitial</td>
<td>5</td>
<td>164</td>
</tr>
<tr>
<td>Testis Leydig Cell</td>
<td>Reproductive</td>
<td>TestisLeydigCell</td>
<td>5</td>
<td>114</td>
</tr>
<tr>
<td>Testis Seminiferous Tubule</td>
<td>Reproductive</td>
<td>TestisSeminiferousTubule</td>
<td>5</td>
<td>100</td>
</tr>
<tr>
<td>CD105 Endothelial</td>
<td>Cardiovascular</td>
<td>BM.CD105.Endothelial</td>
<td>6</td>
<td>426</td>
</tr>
<tr>
<td>CD33 Myeloid</td>
<td>Cardiovascular</td>
<td>BM.CD33.Myeloid</td>
<td>6</td>
<td>472</td>
</tr>
<tr>
<td>CD34</td>
<td>Cardiovascular</td>
<td>BM.CD34.</td>
<td>6</td>
<td>488</td>
</tr>
<tr>
<td>CD71 Early Erythroid</td>
<td>Cardiovascular</td>
<td>BM.CD71.EarlyErythroid</td>
<td>6</td>
<td>461</td>
</tr>
<tr>
<td>Bone Marrow</td>
<td>Cardiovascular</td>
<td>bonemarrow</td>
<td>6</td>
<td>188</td>
</tr>
<tr>
<td>Whole Blood</td>
<td>Cardiovascular</td>
<td>WHOLEBLOOD</td>
<td>6</td>
<td>408</td>
</tr>
<tr>
<td>CD4 T Cells</td>
<td>Lymphatic</td>
<td>CD4.T.cells</td>
<td>6</td>
<td>336</td>
</tr>
<tr>
<td>CD8 T Cells</td>
<td>Lymphatic</td>
<td>CD8.T.cells</td>
<td>6</td>
<td>361</td>
</tr>
<tr>
<td>Lymphnode</td>
<td>Lymphatic</td>
<td>lymphnode</td>
<td>6</td>
<td>154</td>
</tr>
<tr>
<td>BDCA4 Dendritic Cells</td>
<td>Lymphatic</td>
<td>PB.BDCA4.Dendritic_Cells</td>
<td>6</td>
<td>522</td>
</tr>
<tr>
<td>CD14 Monocytes</td>
<td>Lymphatic</td>
<td>PB.CD14.Monocytes</td>
<td>6</td>
<td>355</td>
</tr>
<tr>
<td>CD56 NK Cells</td>
<td>Lymphatic</td>
<td>PB.CD56.NKCells</td>
<td>6</td>
<td>411</td>
</tr>
<tr>
<td>Thymus</td>
<td>Lymphatic</td>
<td>Thymus</td>
<td>6</td>
<td>191</td>
</tr>
<tr>
<td>Tonsil</td>
<td>Lymphatic</td>
<td>Tonsil</td>
<td>6</td>
<td>162</td>
</tr>
<tr>
<td>B Lymphoblasts</td>
<td>Lymphatic</td>
<td>X721_B_lymphoblasts</td>
<td>6</td>
<td>786</td>
</tr>
<tr>
<td>Fetal Liver</td>
<td>Digestive</td>
<td>fetalliver</td>
<td>7</td>
<td>323</td>
</tr>
<tr>
<td>Liver</td>
<td>Digestive</td>
<td>Liver</td>
<td>7</td>
<td>397</td>
</tr>
<tr>
<td>Kidney</td>
<td>Others</td>
<td>Kidney</td>
<td>7</td>
<td>200</td>
</tr>
</tbody>
</table>
Appendix B. **Table S2. Tissue annotations for Figure 4.3.**

Labels: node labeling in Figure 4.3. Color coding: tissues are colored coded according to their categories. File names: data label exactly as in the original dataset. **M:** module number in Figure 4.3. **Sample annotation:** in Arabidopsis, detailed descriptions of sample types are included in this table. **N:** number of genes that are associated with each tissue.

<table>
<thead>
<tr>
<th>Labels</th>
<th>Color coding</th>
<th>File names</th>
<th>M</th>
<th>Sample annotation</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Companion Cell</td>
<td>Root Cells</td>
<td>GSM226527_SUC2_1.CEL</td>
<td>1</td>
<td>Companion cells</td>
<td>456</td>
</tr>
<tr>
<td>Endodermis</td>
<td>Root Cells</td>
<td>scr5_1.CEL</td>
<td>1</td>
<td>Endodermis</td>
<td>297</td>
</tr>
<tr>
<td>Ground Tissues</td>
<td>Root Cells</td>
<td>J0571_1.CEL</td>
<td>1</td>
<td>Ground: endodermins + cortex + quiescent center</td>
<td>233</td>
</tr>
<tr>
<td>Hair Cell</td>
<td>Root Cells</td>
<td>GSM226524_COBL9_1.CEL</td>
<td>1</td>
<td>Hair cell</td>
<td>476</td>
</tr>
<tr>
<td>Lateral Root Cap</td>
<td>Root Cells</td>
<td>LRC_1.CEL</td>
<td>1</td>
<td>Lateral root cap</td>
<td>343</td>
</tr>
<tr>
<td>Lateral Root Primordia</td>
<td>Root Cells</td>
<td>GSM226529_RMI000_1.CEL</td>
<td>1</td>
<td>Lateral root primordia initials</td>
<td>341</td>
</tr>
<tr>
<td>Non-hair Cell</td>
<td>Root Cells</td>
<td>gl2_1.CEL</td>
<td>1</td>
<td>Non hair cell</td>
<td>276</td>
</tr>
<tr>
<td>Pericycle</td>
<td>Root Cells</td>
<td>GSM226528_J2501_1.CEL</td>
<td>1</td>
<td>Pericycle, protoxylem, metaxyylem, phloem</td>
<td>248</td>
</tr>
<tr>
<td>Phloem Pole Pericycle</td>
<td>Root Cells</td>
<td>GSM226522_S17_1.CEL</td>
<td>1</td>
<td>Phloem pole pericycle</td>
<td>400</td>
</tr>
<tr>
<td>Protophloem Metaphloem</td>
<td>Root Cells</td>
<td>GSM226523_S32_1.CEL</td>
<td>1</td>
<td>Protophloem and metaphloem</td>
<td>598</td>
</tr>
<tr>
<td>Protoxylem Metaxylem</td>
<td>Root Cells</td>
<td>GSM226526_S4_1.CEL</td>
<td>1</td>
<td>Protoxylem and 2/3 metaxylem</td>
<td>238</td>
</tr>
<tr>
<td>Stele</td>
<td>Root Cells</td>
<td>wol_1.CEL</td>
<td>1</td>
<td>Stele</td>
<td>336</td>
</tr>
<tr>
<td>Xylem Pole Pericycle</td>
<td>Root Cells</td>
<td>GSM226525_JO121_1.CEL</td>
<td>1</td>
<td>Xylem pole pericycle</td>
<td>243</td>
</tr>
<tr>
<td>7 Days Whole Roots</td>
<td>ATGE_3_A.cel</td>
<td>1 Root, 7 days</td>
<td>249</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8 Days Whole Roots</td>
<td>ATGE_94_A.cel</td>
<td>1 Root, 8 days, MS agar</td>
<td>251</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8 Days* Whole Roots</td>
<td>ATGE_95_A.cel</td>
<td>1 Root, 8 days, MS agar, 1% sucrose</td>
<td>294</td>
<td></td>
<td></td>
</tr>
<tr>
<td>15 Days* Whole Roots</td>
<td>ATGE_93_A.cel</td>
<td>1 Root, 15 days, MS agar, 1% sucrose</td>
<td>296</td>
<td></td>
<td></td>
</tr>
<tr>
<td>17 Days Whole Roots</td>
<td>ATGE_9_A.cel</td>
<td>1 Roots, 17 days</td>
<td>235</td>
<td></td>
<td></td>
</tr>
<tr>
<td>21 Days Whole Roots</td>
<td>ATGE_98_A.cel</td>
<td>1 Root, 21 days, MS agar</td>
<td>239</td>
<td></td>
<td></td>
</tr>
<tr>
<td>21 Days* Whole Roots</td>
<td>ATGE_99_A.cel</td>
<td>1 Root, 21 days, MS agar, 1% sucrose</td>
<td>275</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stage I Whole Roots</td>
<td>StageI_1.CEL</td>
<td>1 Stage I</td>
<td>195</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stage</td>
<td>Whole Roots</td>
<td>Stage CEL</td>
<td>Count</td>
<td>Part</td>
<td>Description</td>
</tr>
<tr>
<td>-------</td>
<td>-------------</td>
<td>-----------</td>
<td>-------</td>
<td>------</td>
<td>-------------</td>
</tr>
<tr>
<td>Stage II</td>
<td>Whole Roots</td>
<td>StageII_1.CEL</td>
<td>1</td>
<td>Stage II</td>
<td>211</td>
</tr>
<tr>
<td>Stage III</td>
<td>Whole Roots</td>
<td>StageIII_2.CEL</td>
<td>1</td>
<td>Stage III</td>
<td>422</td>
</tr>
<tr>
<td>Hypocotyl</td>
<td>Others</td>
<td>ATGE_2_A.cel</td>
<td>1</td>
<td>Hypocotyl</td>
<td>95</td>
</tr>
<tr>
<td>Stage 8</td>
<td>Seeds</td>
<td>ATGE_82_A.cel</td>
<td>2</td>
<td>Seed stage 8</td>
<td>223</td>
</tr>
<tr>
<td>Stage 10</td>
<td>Seeds</td>
<td>ATGE_84_A.cel</td>
<td>2</td>
<td>Seed stage 10</td>
<td>198</td>
</tr>
<tr>
<td>Stage 9</td>
<td>Seeds</td>
<td>ATGE_83_A.cel</td>
<td>2</td>
<td>Seed stage 9</td>
<td>195</td>
</tr>
<tr>
<td>Stage 7</td>
<td>Seeds</td>
<td>ATGE_81_A.cel</td>
<td>2</td>
<td>Seed stage 7</td>
<td>177</td>
</tr>
<tr>
<td>Stage 6</td>
<td>Seeds</td>
<td>ATGE_79_A.cel</td>
<td>2</td>
<td>Seed, stage 6</td>
<td>164</td>
</tr>
<tr>
<td>21 Days*</td>
<td>Seedling Green Parts</td>
<td>ATGE_101_A.cel</td>
<td>3</td>
<td>Seedling, green parts, MS agar, 1% sucrose</td>
<td>40</td>
</tr>
<tr>
<td>21 Days</td>
<td>Seedling Green Parts</td>
<td>ATGE_100_A.cel</td>
<td>3</td>
<td>Seedling, green parts, MS agar</td>
<td>32</td>
</tr>
<tr>
<td>8 Days*</td>
<td>Seedling Green Parts</td>
<td>ATGE_97_A.cel</td>
<td>3</td>
<td>Seedling green parts, 8 days, MS agar, 1% sucrose</td>
<td>47</td>
</tr>
<tr>
<td>8 Days</td>
<td>Seedling Green Parts</td>
<td>ATGE_96_A.cel</td>
<td>3</td>
<td>Seedling green parts, 8 days, MS agar</td>
<td>52</td>
</tr>
<tr>
<td>1st Internode</td>
<td>Stems</td>
<td>ATGE_28_A2.cel</td>
<td>4</td>
<td>Stem 1st node</td>
<td>118</td>
</tr>
<tr>
<td>2nd Internode</td>
<td>Stems</td>
<td>ATGE_27_A.cel</td>
<td>4</td>
<td>Stem 2nd node</td>
<td>115</td>
</tr>
<tr>
<td>Vegetative Shoot Apex</td>
<td>ATGE_6_A.cel</td>
<td>5</td>
<td>Shoot apex, vegetative, 7 days</td>
<td>30</td>
<td></td>
</tr>
<tr>
<td>Vegetative Young Leaves Shoot Apex</td>
<td>ATGE_4_A.cel</td>
<td>5</td>
<td>Shoot apex, vegetative + young leaves</td>
<td>16</td>
<td></td>
</tr>
<tr>
<td>7 Days</td>
<td>Seedling Green Parts</td>
<td>ATGE_7_A2.cel</td>
<td>5</td>
<td>Seedling green parts, 7 days</td>
<td>20</td>
</tr>
<tr>
<td>7 Days</td>
<td>Rosette Leaves</td>
<td>ATGE_87_A.cel</td>
<td>5</td>
<td>Vegetative rosette, 7 days</td>
<td>23</td>
</tr>
<tr>
<td>14 Days</td>
<td>Rosette Leaves</td>
<td>ATGE_89_A.cel</td>
<td>5</td>
<td>Vegetative rosette, 14 days</td>
<td>18</td>
</tr>
<tr>
<td>21 Days</td>
<td>Rosette Leaves</td>
<td>ATGE_90_A.cel</td>
<td>5</td>
<td>Vegetative rosette, 21 days</td>
<td>17</td>
</tr>
<tr>
<td>21 Days</td>
<td>Rosette Leaves</td>
<td>ATGE_22_A.cel</td>
<td>6</td>
<td>Entire rosette, developmental drift, 21 days</td>
<td>12</td>
</tr>
<tr>
<td>22 Days</td>
<td>Rosette Leaves</td>
<td>ATGE_23_A.cel</td>
<td>6</td>
<td>Entire rosette, developmental drift, 22 days</td>
<td>12</td>
</tr>
<tr>
<td>23 Days</td>
<td>Rosette Leaves</td>
<td>ATGE_24_A.cel</td>
<td>6</td>
<td>Entire rosette, developmental drift, 23 days</td>
<td>23</td>
</tr>
<tr>
<td>No. 2</td>
<td>Rosette Leaves</td>
<td>ATGE_12_A.cel</td>
<td>6</td>
<td>Rosette leaf 2</td>
<td>35</td>
</tr>
<tr>
<td>No.</td>
<td>Rosette Leaves</td>
<td>ATGE_{13} A.cel</td>
<td>6</td>
<td>Rosette leaf 4</td>
<td>22</td>
</tr>
<tr>
<td>------</td>
<td>------------------------</td>
<td>----------------</td>
<td>----</td>
<td>----------------</td>
<td>----</td>
</tr>
<tr>
<td>No.</td>
<td>Rosette Leaves</td>
<td>ATGE_{14} A.cel</td>
<td>6</td>
<td>Rosette leaf 6</td>
<td>16</td>
</tr>
<tr>
<td>No.</td>
<td>Distal Half Rosette</td>
<td>ATGE_{21} A.cel</td>
<td>6</td>
<td>Leaf 7, distal half</td>
<td>17</td>
</tr>
<tr>
<td>No.</td>
<td>Proximal Half Rosette</td>
<td>ATGE_{20} A.cel</td>
<td>6</td>
<td>Leaf 7, proximal half</td>
<td>12</td>
</tr>
<tr>
<td>No.</td>
<td>Rosette Leaves</td>
<td>ATGE_{15} A.cel</td>
<td>6</td>
<td>Rosette leaf 8</td>
<td>9</td>
</tr>
<tr>
<td>No.</td>
<td>Rosette Leaves</td>
<td>ATGE_{16} A.cel</td>
<td>6</td>
<td>Rosette leaf 10</td>
<td>9</td>
</tr>
<tr>
<td>No.</td>
<td>Rosette Leaves</td>
<td>ATGE_{17} A.cel</td>
<td>6</td>
<td>Rosette leaf 12</td>
<td>8</td>
</tr>
<tr>
<td>Cau</td>
<td>Rosette Leaves</td>
<td>ATGE_{26} A.cel</td>
<td>6</td>
<td>Cauline leaves</td>
<td>38</td>
</tr>
<tr>
<td>Sen</td>
<td>Rosette Leaves</td>
<td>ATGE_{25} A.cel</td>
<td>6</td>
<td>Senescing leaves</td>
<td>125</td>
</tr>
<tr>
<td>Cot</td>
<td>Others</td>
<td>ATGE_{1} A.cel</td>
<td>6</td>
<td>Cotyledons 7 days</td>
<td>20</td>
</tr>
<tr>
<td>Before Bolting</td>
<td>Shoot Apex</td>
<td>ATGE_{8} A.cel</td>
<td>7</td>
<td>Shoot apex, transition, before bolting, 14 days</td>
<td>40</td>
</tr>
<tr>
<td>Inflorescence</td>
<td>Shoot Apex</td>
<td>ATGE_{29} A2.cel</td>
<td>7</td>
<td>Shoot apex, inflorescence (after bolting)</td>
<td>35</td>
</tr>
<tr>
<td>Stage 9</td>
<td>Whole Flower</td>
<td>ATGE_{31} A2.cel</td>
<td>8</td>
<td>Flowers stage 9</td>
<td>201</td>
</tr>
<tr>
<td>Stage 10/11</td>
<td>Whole Flower</td>
<td>ATGE_{32} A2.cel</td>
<td>8</td>
<td>Flowers stage 10/11</td>
<td>184</td>
</tr>
<tr>
<td>Stage 12</td>
<td>Whole Flower</td>
<td>ATGE_{33} A.cel</td>
<td>8</td>
<td>Flower stage 12</td>
<td>451</td>
</tr>
<tr>
<td>Stage 15</td>
<td>Whole Flower</td>
<td>ATGE_{39} A.cel</td>
<td>8</td>
<td>Flower stage 15</td>
<td>390</td>
</tr>
<tr>
<td>28 Days</td>
<td>Whole Flower</td>
<td>ATGE_{92} A.cel</td>
<td>8</td>
<td>Flower, 28 days, soil</td>
<td>273</td>
</tr>
<tr>
<td>Carpels Stage 12</td>
<td>Floral Organs</td>
<td>ATGE_{37} A.cel</td>
<td>8</td>
<td>Flower stage 12, carpels</td>
<td>85</td>
</tr>
<tr>
<td>Carpels Stage 15</td>
<td>Floral Organs</td>
<td>ATGE_{45} A.cel</td>
<td>8</td>
<td>Flower stage 15, carpels</td>
<td>251</td>
</tr>
<tr>
<td>Petals Stage 12</td>
<td>Floral Organs</td>
<td>ATGE_{35} A.cel</td>
<td>8</td>
<td>Flower stage 12, petals</td>
<td>229</td>
</tr>
<tr>
<td>Petals Stage 15</td>
<td>Floral Organs</td>
<td>ATGE_{42} B.cel</td>
<td>8</td>
<td>Flower stage 15, petals</td>
<td>261</td>
</tr>
<tr>
<td>Sepals Stage 12</td>
<td>Floral Organs</td>
<td>ATGE_{34} A.cel</td>
<td>8</td>
<td>Flower stage 12, sepals</td>
<td>144</td>
</tr>
<tr>
<td>Sepals Stage 15</td>
<td>Floral Organs</td>
<td>ATGE_{41} A.cel</td>
<td>8</td>
<td>Flower stage 15, sepals</td>
<td>291</td>
</tr>
<tr>
<td>Stamens Stage 12</td>
<td>Floral Organs</td>
<td>ATGE_{36} A.cel</td>
<td>8</td>
<td>Flower stage 12, stamen</td>
<td>967</td>
</tr>
<tr>
<td>Stamens Stage 15</td>
<td>Floral Organs</td>
<td>ATGE_{43} A.cel</td>
<td>8</td>
<td>Flower stage 15, stamen</td>
<td>686</td>
</tr>
<tr>
<td>Stage 3</td>
<td>Siliques</td>
<td>ATGE_{76} A.cel</td>
<td>8</td>
<td>Silique, stage 3</td>
<td>379</td>
</tr>
<tr>
<td>Stage 4</td>
<td>Siliques</td>
<td>ATGE_77_D.cel</td>
<td>8</td>
<td>Siliques, stage 4</td>
<td>280</td>
</tr>
<tr>
<td>--------</td>
<td>----------</td>
<td>---------------</td>
<td>---</td>
<td>-------------------</td>
<td>-----</td>
</tr>
<tr>
<td>Stage 5</td>
<td>Siliques</td>
<td>ATGE_78_D.cel</td>
<td>8</td>
<td>Siliques, stage 5</td>
<td>227</td>
</tr>
</tbody>
</table>
Appendix B. Table S3. HKGs with known functional roles in guard cells.

<table>
<thead>
<tr>
<th>Accession</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AT4G34000</td>
<td>ABF3 (ABSCISIC ACID RESPONSIVE ELEMENTS-BINDING FACTOR 3)</td>
</tr>
<tr>
<td>AT4G26080</td>
<td>ABI1 (ABA INSENSITIVE 1)</td>
</tr>
<tr>
<td>AT5G57050</td>
<td>ABI2 (ABA INSENSITIVE 2)</td>
</tr>
<tr>
<td>AT3G61640</td>
<td>AGP20</td>
</tr>
<tr>
<td>AT2G18960</td>
<td>AHA1/OST2 (ARABIDOPSIS H⁺ ATPASE 1)</td>
</tr>
<tr>
<td>AT3G53720</td>
<td>ATCHX20 (CATION/H⁺ EXCHANGER 20)</td>
</tr>
<tr>
<td>AT4G17615</td>
<td>CBL1 (CALCINEURIN B-LIKE PROTEIN 1)</td>
</tr>
<tr>
<td>AT5G05410</td>
<td>DREB2A</td>
</tr>
<tr>
<td>AT3G11020</td>
<td>DREB2B (DRE/CRT-BINDING PROTEIN 2B)</td>
</tr>
<tr>
<td>AT2G20875</td>
<td>EPF1 (EPIDERMAL PATTERNING FACTOR 1)</td>
</tr>
<tr>
<td>AT1G66340</td>
<td>ETR1 (ETHYLENE RESPONSE 1)</td>
</tr>
<tr>
<td>AT3G24140</td>
<td>FMA (FAMA)</td>
</tr>
<tr>
<td>AT5G37500</td>
<td>GORK (GATED OUTWARDLY-RECTIFYING K⁺ CHANNEL)</td>
</tr>
<tr>
<td>AT1G62400</td>
<td>HT1 (HIGH LEAF TEMPERATURE 1)</td>
</tr>
<tr>
<td>AT3G26744</td>
<td>ICE1 (INDUCER OF CBF EXPRESSION 1)</td>
</tr>
<tr>
<td>AT5G46240</td>
<td>KAT1 (POTASSIUM CHANNEL IN ARABIDOPSIS THALIANA 1)</td>
</tr>
<tr>
<td>AT4G18290</td>
<td>KAT2 (POTASSIUM CHANNEL IN ARABIDOPSIS THALIANA 2)</td>
</tr>
<tr>
<td>AT1G08810</td>
<td>MYB60 (MYB DOMAIN PROTEIN 60)</td>
</tr>
<tr>
<td>AT1G78390</td>
<td>NCED9 (NINE-CIS-EPOXYCAROTENOID DIOXYGENASE 9)</td>
</tr>
<tr>
<td>AT4G33950</td>
<td>OST1 (OPEN STOMATA 1)</td>
</tr>
<tr>
<td>AT1G12480</td>
<td>OZS1/SLAC1 (OZONE-SENSITIVE 1)</td>
</tr>
<tr>
<td>AT2G29940</td>
<td>PDR3 (PLEIOTROPIC DRUG RESISTANCE 3)</td>
</tr>
<tr>
<td>AT3G11820</td>
<td>SYP121 (SYNTAXIN OF PLANTS 121)</td>
</tr>
<tr>
<td>AT1G80080</td>
<td>TMM (TOO MANY MOUTHS)</td>
</tr>
</tbody>
</table>
### Appendix B. Table S4.

Gene Ontology (GO) analysis of leptokurtic genes in guard cells and neighboring tissues. **GO id:** Gene Ontology id for each enriched GO term. **GO annotation:** Functional annotation for each GO id. Other columns list enrichment p values for each GO term as calculated by eliminate count method using topGO software. Sample names for each column label can be found in Appendix B. Table S2.

<table>
<thead>
<tr>
<th>GO id</th>
<th>GO annotation</th>
<th>GC</th>
<th>gl2</th>
<th>COBL9</th>
<th>LRC</th>
<th>ATGE82</th>
<th>ATGE83</th>
<th>ATGE84</th>
<th>ATGE25</th>
<th>ATGE41</th>
<th>src5</th>
<th>wol</th>
</tr>
</thead>
<tbody>
<tr>
<td>GO:0004365</td>
<td>glyceraldehyde-3-phosphate...</td>
<td>1.E-06</td>
<td>3.E-06</td>
<td>4.E-04</td>
<td>2.E-04</td>
<td>1.E+00</td>
<td>1.E+00</td>
<td>1.E+00</td>
<td>1.E+00</td>
<td>1.E+00</td>
<td>2.E-04</td>
<td>2.E-04</td>
</tr>
<tr>
<td>Gene ID</td>
<td>Description</td>
<td>E-values</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>--------</td>
<td>------------------------------------------------------------------------------</td>
<td>----------</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GO:0005762</td>
<td>mitochondrial large ribosome remodeling</td>
<td>1.0E-04</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GO:0009753</td>
<td>response to jasmonic acid</td>
<td>1.0E-04</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GO:0010294</td>
<td>abscisic acid glucosyltransfer</td>
<td>5.0E-04</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GO:0009751</td>
<td>response to salicylic acid</td>
<td>8.0E-04</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GO:0043295</td>
<td>glutathione binding</td>
<td>1.0E-03</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GO:0055114</td>
<td>oxidation reduction</td>
<td>2.0E-03</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GO:0005740</td>
<td>mitochondrial envelope</td>
<td>1.0E-01</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GO:0009636</td>
<td>response to toxin</td>
<td>2.0E-03</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GO:0012505</td>
<td>endomembrane system</td>
<td>2.0E-03</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GO:0009409</td>
<td>response to cold</td>
<td>3.0E-03</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GO:0000038</td>
<td>very-long-chain fatty acid regulation</td>
<td>3.0E-03</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GO:0051302</td>
<td>regulation of cell division</td>
<td>3.0E-03</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GO:0006096</td>
<td>Glycolysis</td>
<td>3.0E-03</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GO:0004707</td>
<td>MAP kinase activity</td>
<td>3.0E-03</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GO:0006984</td>
<td>ER-nuclear signaling pathway response to unfolded</td>
<td>3.0E-03</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GO:0034620</td>
<td>response to unfolded</td>
<td>3.0E-03</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GO:0034976</td>
<td>response to ER stress</td>
<td>3.0E-03</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GO:0008287</td>
<td>protein S/T phosphatase</td>
<td>4.0E-03</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Appendix B. Table S5. Primers used for real time PCR analysis of GC-abundant genes and for genotyping of knock-out lines.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Forward primer</th>
<th>Reverse primer</th>
</tr>
</thead>
<tbody>
<tr>
<td>At2g22320</td>
<td>GCGGGAGTGGATTATCAAGA</td>
<td>AATTTCCCAACCAGAACGTCTC</td>
</tr>
<tr>
<td>At1g64010</td>
<td>TGTGGAGAAAAAGCGCATCAA</td>
<td>AAAGCCTCCAAACAAGAGCAG</td>
</tr>
<tr>
<td>At2g21080</td>
<td>AATGTGTCACCCTCACTGGGA</td>
<td>TCAGGATCCACATCACCAGA</td>
</tr>
<tr>
<td>At3g51760</td>
<td>TCTAGCAGCCCTAGACATCAG</td>
<td>AACTTGAGGTGCAAGTGAGCAG</td>
</tr>
<tr>
<td>At4g16820</td>
<td>CGTCACTCCTTCTCGTTCACCT</td>
<td>AGGCATTGAAACCAAGAGATGG</td>
</tr>
<tr>
<td>At1g11100</td>
<td>GTCCCGCGACAGTAATGAA</td>
<td>GGTCAGCATCTTTGTCATC</td>
</tr>
<tr>
<td>At4g28460</td>
<td>GTGGTTGTGATGGTGTCGTT</td>
<td>CCAACGGCTGAACACCAAGA</td>
</tr>
<tr>
<td>At1g50400</td>
<td>TCCGTCAGGGCTAATGCTC TCC</td>
<td>CTCAACCCCAACCAGCACGTGT</td>
</tr>
<tr>
<td>At2g19180</td>
<td>CTCTTTCTTG ACTCTCTCTC</td>
<td>GAATTATGAAACGATGATCT</td>
</tr>
<tr>
<td>At5g35320</td>
<td>GCCGGTTAAGTCATGAGAGAG</td>
<td>TTCCAGCAAACAACGCTCTG</td>
</tr>
<tr>
<td>At4g11330</td>
<td>ATCCACCTGATGGAGCGAGTCT</td>
<td>GTGGAACAAACAGCTGCTG</td>
</tr>
<tr>
<td>At5g60410</td>
<td>TTTGGGTTACAGTGAGCAGA</td>
<td>CTAACAGGGACACATTCGTG</td>
</tr>
<tr>
<td>At1g20880</td>
<td>CACTGAGTCCATCCAAGGTT</td>
<td>CAGGACATGGGCTATACCT</td>
</tr>
<tr>
<td>At2g27200</td>
<td>AGCAACTTCTTGGGAGAGTCTT</td>
<td>GCTTGGCTGAGCTCGAGAACAG</td>
</tr>
<tr>
<td>At2g18790</td>
<td>AGAATTGCTGCTGAAGTAGTCTC</td>
<td>CCGTTCTGATTCTCGGATGTC</td>
</tr>
<tr>
<td>(PhyB)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>At1g11100-1</td>
<td>CCAAGGAAAGAACGTGATTTG</td>
<td>ACAGATAACCTCAAATCCTCC</td>
</tr>
<tr>
<td>At1g11100-2</td>
<td>CGGTTGACATTCTCAATCCTC</td>
<td>ACAGATAACCTCAAATCCTCC</td>
</tr>
<tr>
<td>At2g21080-1</td>
<td>TGGATACCTTCTCCTCTGTG</td>
<td>AGTCAGTGGTGTACCTTCGGT</td>
</tr>
<tr>
<td>At2g21080-2</td>
<td>GGAAGATATGTAATCCCTGCTGT</td>
<td>TCCAGACTGTTGAAACATGGGA</td>
</tr>
<tr>
<td>At1g50400-1;</td>
<td>GAGGAACCTCAACCGTGAAGC</td>
<td>CAAGATGGACTGACGCGACAACA</td>
</tr>
<tr>
<td>At1g50400-2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(siz 1-2)</td>
<td>GCTAGGCTTGAACCGCATCCAG</td>
<td>AAAGAGAGAGAGTAGGAGCGAAGG</td>
</tr>
<tr>
<td>(siz 1-3)</td>
<td>GAGCTGAGGATCCTCGTGTGTGT</td>
<td>CAGCAGAGATGGAAGCATTGTG</td>
</tr>
</tbody>
</table>

Primers used for genotyping of knock-out lines.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Forward primer</th>
<th>Reverse primer</th>
</tr>
</thead>
<tbody>
<tr>
<td>At1g11100-1</td>
<td>CCAAGGAAAGAACGTGATTTG</td>
<td>ACAGATAACCTCAAATCCTCC</td>
</tr>
<tr>
<td>At1g11100-2</td>
<td>CGGTTGACATTCTCAATCCTC</td>
<td>ACAGATAACCTCAAATCCTCC</td>
</tr>
<tr>
<td>At2g21080-1</td>
<td>TGGATACCTTCTCCTCTGTG</td>
<td>AGTCAGTGGTGTACCTTCGGT</td>
</tr>
<tr>
<td>At2g21080-2</td>
<td>GGAAGATATGTAATCCCTGCTGT</td>
<td>TCCAGACTGTTGAAACATGGGA</td>
</tr>
<tr>
<td>At1g50400-1;</td>
<td>GAGGAACCTCAACCGTGAAGC</td>
<td>CAAGATGGACTGACGCGACAACA</td>
</tr>
<tr>
<td>At1g50400-2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(siz 1-2)</td>
<td>GCTAGGCTTGAACCGCATCCAG</td>
<td>AAAGAGAGAGAGTAGGAGCGAAGG</td>
</tr>
<tr>
<td>(siz 1-3)</td>
<td>GAGCTGAGGATCCTCGTGTGTGT</td>
<td>CAGCAGAGATGGAAGCATTGTG</td>
</tr>
</tbody>
</table>
Works Cited


209


233. Du, Z., K. Aghoram, and W.H. Outlaw, Jr., *In vivo phosphorylation of phosphoenolpyruvate carboxylase in guard cells of Vicia faba L. is enhanced by*


215


Vita

Song Li

Education:

2006-Present Ph.D. candidate, expected graduation date: Dec 2010, Bioinformatics and Genomics, Integrative Biosciences (IBIOS), The Huck Institutes of The Life Sciences, Pennsylvania State University

2002-2006 Ph.D. candidate, Ecological and Molecular Plant Physiology, Integrative Biosciences (IBIOS), The Huck Institutes of The Life Sciences, Pennsylvania State University

1998-2002 B.S. in Biotechnology, Peking University

Publications:

Journal Publication:

Song Li, Sona Pandey, Liza Wilson, Zhixin Zhao, Timothy E. Gookin, Sarah M. Assmann. High kurtosis gene expression and gene-centric tissue networks reveal new functional systems principles of Arabidopsis transcriptomes (Submitted).


Song Li, Sarah Assmann, Réka Albert. Predicting essential components of signal transduction networks. PLOS Biology. 2006 Volume 4, No.10, e312,

Genji Qin, Dingming Kang, Yiyu Dong, Yuping Shen, Li Zhang, Xiaohui Deng, Yao Zhang, Song Li, Nan Chen, Weiran Niu, Cong Chen, Peicheng Liu, Hongya Gu, Xingwang Deng, Li-Jia Qu and Zhangliang Chen, Obtaining and analysis of flanking sequences from T-DNA transformants in Arabidopsis, Plant Science, 2003,165: 941-949

Book chapter:


Teaching Experience:


2001-2002 Supervised undergraduate student research (lab) in Peking University-Yale University joint center (Beijing) of plant molecular biology and biotechnology