# Boolean modeling of transcriptome data reveals novel modes of heterotrimeric G-protein action

**Sona Pandey[1,3], Rui-Sheng Wang[2,3], Liza Wilson[1], Song Li[1], Zhixin Zhao[1], Timothy E Gookin[1], Sarah M Assmann[1] and Réka Albert[2,]***

[1] Department of Biology, Pennsylvania State University, University Park, PA, USA and [2] Department of Physics, Pennsylvania State University, University Park, PA, USA
[3] These authors contributed equally to this work
* Corresponding author. Department of Physics, Pennsylvania State University, 104 Davey Laboratory, University Park, PA 16802, USA.
Tel.: + 1 814 865 6123; Fax: + 1 814 865 3604; E-mail: ralbert@phys.psu.edu

**Heterotrimeric G-proteins mediate crucial and diverse signaling pathways in eukaryotes. Here, we generate and analyze microarray data from guard cells and leaves of G-protein subunit mutants of the model plant *Arabidopsis thaliana,* with or without treatment with the stress hormone, abscisic acid. Although G-protein control of the transcriptome has received little attention to date in any system, transcriptome analysis allows us to search for potentially uncommon yet significant signaling mechanisms. We describe the theoretical Boolean mechanisms of G-protein × hormone regulation, and then apply a pattern matching approach to associate gene expression profiles with Boolean models. We find that (1) classical mechanisms of G-protein signaling are well represented. Conversely, some theoretical regulatory modes of the G-protein are not supported; (2) a new mechanism of G-protein signaling is revealed, in which Gβ regulates gene expression identically in the presence or absence of Gα; (3) guard cells and leaves favor different G-protein modes in transcriptome regulation, supporting system specificity of G-protein signaling. Our method holds significant promise for analyzing analogous 'switch-like' signal transduction events in any organism.**

## Introduction

Heterotrimeric G-proteins, composed of α, β, and γ subunits, participate in a wide range of signaling pathways in eukaryotes (Morris and Malbon, 1999; Jones and Assmann, 2004; Perfus-Barbeoch *et al*, 2004). According to the typical, mammalian paradigm, in its inactive state, the G-protein exists as an associated heterotrimer. G-protein signaling begins with ligand binding that results in a conformational change in a G-protein-coupled receptor (GPCR). Once activated by the GPCR, the Gα protein, which possesses a GDP/GTP-nucleotide-binding site and GTP-hydrolase activity, changes its form to a structure that allows exchange of GDP for GTP. The GTP-bound Gα separates from the associated Gβγ dimer and the freed Gα and Gβγ proteins can then interact with downstream effector molecules, alone or in combination, to transduce the signal. Subsequent to signal propagation, the intrinsic GTPase activity of Gα eventually results in hydrolysis of bound GTP to GDP, which inactivates Gα and allows its re-association with the Gβγ dimer to reform the inactive G-protein complex.

There are several classical routes for signal propagation through heterotrimeric G-proteins that have been categorized in mammalian systems (Morris and Malbon, 1999; Marrari *et al*, 2007; Oldham and Hamm, 2008; Dupre *et al*, 2009). On the basis of the requirement for both Gα and Gβγ subunits or for only the Gβγ subunit to propagate the signal, we group these routes into classical I and classical II categories, respectively. In one mechanism of the classical I type (designated here as classical Ia), on GPCR activation, freed Gα and Gβγ both interact with downstream effectors, either additively or synergistically, to elicit the downstream response. In a second mechanism (designated here as classical Ib), Gα but not Gβγ interacts with downstream effectors; in this case, there still is a requirement for the Gβγ dimer to facilitate correct targeting of Gα to the plasma membrane and coupling of Gα with the relevant GPCR (Marrari *et al*, 2007). In the classical II route, well established in mammalian and yeast

cells, it is solely the Gβγ dimer that interacts with downstream effectors; in this case, sequestration of Gβγ within the heterotrimer prevents signal propagation.

Each one of these classical mechanisms leads to explicit predictions of how genetic knockout of either Gα or Gβ would affect production of the downstream phenotype. In classical Ia, knockout of either Gα or Gβ would lead to an impaired response. Either there could be a stronger impairment in the case of a double knockout of Gα and Gβ, or a single knockout could suffice to eliminate the response, if the single knockout reduced the signal to below a requisite activation threshold. In classical Ib, knockout of Gβ would prevent correct Gα localization and Gα-GPCR coupling, whereas knockout of Gα would eliminate Gα-effector coupling. Thus, each single knockout as well as a double knockout of Gα and Gβ would eliminate the response. In the classical II mechanism, knockout of Gα would free Gβ(γ) from sequestration, thereby eliciting the phenotype even in the absence of agonist activation of the GPCR; conversely, knockout of Gβ would defacto eliminate the possibility of eliciting the phenotype, even in the presence of the appropriate ligand.

Direct genetic tests of the above predictions are complicated in mammalian systems, due to the abundance of G-protein subunits (e.g. 23 Gα subunits (including splice variants), 5 Gβ subunits, and 12 Gγ subunits encoded in the human genome (McCudden *et al*, 2005)) and their potential for combinatorial associations and partially redundant function. However, the presence of only one canonical Gα protein (GPA1), one Gβ protein (AGB1), and two identified Gγ proteins (AGG1 and AGG2) in the model plant *Arabidopsis thaliana* (Jones and Assmann, 2004; Assmann, 2005), and the availability of viable genetic knockout mutants lacking Gα (*gpa1*), Gβ (*agb1*), or both (*agb1 gpa1* double) (Lease *et al*, 2001; Jones *et al*, 2003; Ullah *et al*, 2003; Chen *et al*, 2004), offers an excellent opportunity to experimentally test these predictions. Although biochemical analyses of plant G-protein receptor and subunit-effector coupling are just beginning, plant G-proteins, like those of animals, have been shown to participate in multiple signaling and developmental processes, and phenotypic analysis of G-protein mutants suggests that the above classical mechanisms also exist in plants. For example, some phenotypes, such as rounded rosette leaves, are exhibited similarly by both *gpa1* and *agb1* knockout mutants (Assmann, 2005), consistent with the classical Ia and Ib mechanisms. Other phenotypes are opposite in *gpa1* and *agb1* mutants, supporting a classical II mechanism. For example, *agb1* mutants exhibit increased numbers of lateral roots compared with wild type, whereas *gpa1* mutants show decreased lateral root production (Chen *et al*, 2006), and *agb1* mutants exhibit impaired resistance to some plant pathogens, whereas *gpa1* mutants exhibit enhanced resistance (Trusov *et al*, 2006).

In addition to the two classical mechanisms discussed above, a few non-classical G-protein regulatory modes have also been implicated in some systems, for example signaling by the intact heterotrimer in yeast, the possibility of varying extents of heterotrimer dissociation in mammalian cells (Klein *et al*, 2000; Frank *et al*, 2005; Digby *et al*, 2008), and a suggestion that Gα in *Arabidopsis* exists primarily in a GTP-bound, dissociated state (Johnston *et al*, 2007; Temple and Jones, 2007). Observations such as these lead to a fundamental question, namely, which of all the theoretical regulatory modes in G-protein signaling are biologically possible, exemplifying a more general question of how we can best model the effects of switch-like signaling mechanisms that have multiple active states. It is these two questions that are addressed here.

To facilitate the discovery of non-classical mechanisms, which arguably occur more rarely than well-established classical mechanisms, here we generate microarray data from wild-type, *gpa1*, *agb1*, and *agb1 gpa1* mutant plants and use transcriptome analysis, in which thousands of outputs (i.e. levels of individual transcripts) can be monitored simultaneously. To assess cell/tissue specificity of G-protein signaling mechanisms, we perform transcriptome analysis in two types of samples, stomatal guard cells and rosette leaves.

We also assay these transcriptomes in the presence or absence of the phytohormone abscisic acid (ABA), a major plant hormone that both inhibits growth and promotes tolerance of abiotic stresses such as drought, salinity, and cold (Leung and Giraudat, 1998; Finkelstein *et al*, 2002; Acharya and Assmann, 2009). Although a few dozen candidate plant GPCRs with predicted 7TM structure have been computationally identified (Moriyama *et al*, 2006; Gookin *et al*, 2008), and several of these have been shown experimentally to interact with GPA1 (Gookin *et al*, 2008), to date none of these proteins has an identified ligand. We chose ABA as a variable because ABA signaling is known to interact with heterotrimeric G-protein signaling in both developmental and stress responses in a complex manner (Wang *et al*, 2001; Pandey *et al*, 2006, 2009; Fan *et al*, 2008). For example, ABA inhibition of stomatal opening, which promotes water conservation under stress conditions by reducing water vapor efflux through microscopic stomatal pores at the leaf surface, is impaired in *gpa1* and *agb1* single mutants as well as *agb1 gpa1* double mutants, exemplifying ABA hyposensitivity of guard cell processes (Wang *et al*, 2001; Coursol *et al*, 2003; Fan *et al*, 2008). By contrast, seed germination and post-germination seedling development are hypersensitive to inhibition by ABA in G-protein complex mutants (Pandey *et al*, 2006). These experimental observations suggest G-proteins as one of the components of ABA signaling, but to date no systematic study has been conducted to define the regulatory modes of a G-protein or the co-regulatory modes of a G-protein and a hormone. Further, regulation of gene expression in G-protein complex mutants has rarely received genome-wide analysis in either plant or mammalian systems (Ullah *et al*, 2003; Okamoto *et al*, 2009), and there is only one replicated study of the guard cell transcriptome (Leonhardt *et al*, 2004).

As there is no evidence for ABA or heterotrimeric G-proteins directly interacting with chromatin or DNA to influence gene expression, in this study we posit one or more mediators acting downstream of these inputs to control transcript levels. For simplicity, we use the minimum number of mediators necessary to explain the regulatory modes. In reality, the number of components involved may be more. In the two simplest cases, transcript levels would be (1) controlled by ABA independently of the G-protein, or, conversely (2) controlled by the G-protein independently of ABA, that is on activation by a signal other than ABA (Figure 1A and B). The simplest case of a combined G-protein–ABA effect on a gene's
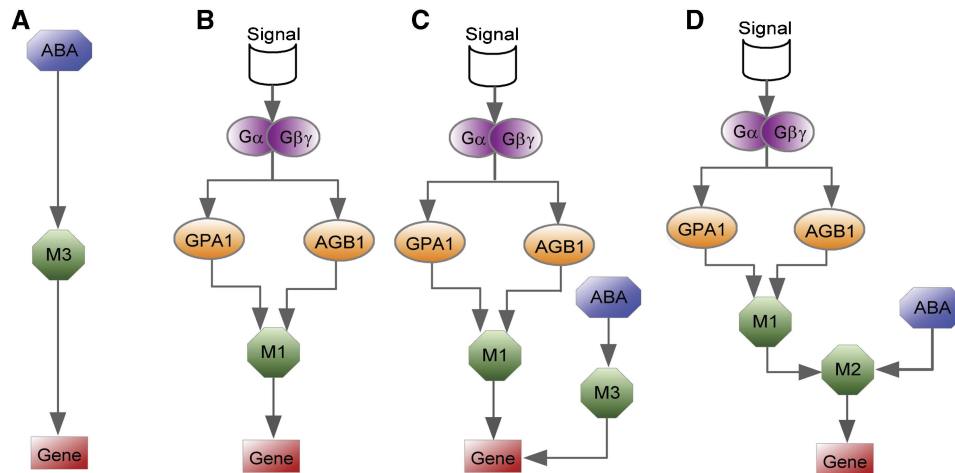
**Figure 1** Illustration of possible regulatory modes of the heterotrimeric G-protein and ABA. The activation of the G-protein and its dissociation into its subunits is illustrated as an arrow activated by a signal. The effector M1 can be regulated by either GPA1, AGB1, or both G-protein subunits; for simplicity, these possibilities are not broken out into separate diagrams. The effector M2 is co-regulated by the G-protein and ABA. The effector M3 is regulated by ABA, independent of the G-protein. The regulatory effect on mediators and genes can be activation or inhibition. The illustrations depict the minimum number of mediators necessary to explain the regulatory modes. In reality, the number of components involved may be more. (**A**) Transcript levels are controlled by ABA independently of the G-protein. (**B**) Transcript levels are controlled by the G-protein independently of ABA. (**C**) Transcript levels are additively controlled by the G-protein and ABA. (**D**) Transcript levels are combinatorially controlled by the G-protein and ABA.

expression would be that in addition to the G-protein regulation, ABA has an additive G-protein-independent regulatory effect on the gene through another signaling mediator, as shown in Figure 1C. It is also possible that after the G-protein is activated by a signal other than ABA, the G-protein subunit(s) and ABA combinatorially co-regulate a mediator, which then directly or indirectly regulates the gene, as shown in Figure 1D. There is also the possibility that ABA serves as a signal that activates the G-protein. Combined with the regulatory modes illustrated in Figure 1, a total of nine possible G-protein and/or ABA signaling pathways can be distinguished (Supplementary Figure S1). On the basis of nine G-protein/ABA signaling pathways and the combinatorial relationships between G-protein subunits, we use a Boolean modeling approach to systematically enumerate all possible theoretical regulatory modes of the G-protein and co-regulatory modes of the G-protein and ABA (a total of 142 regulatory modes). In this study, we determine which signaling pathways and regulatory modes are experimentally supported by transcriptome data.

The Boolean modeling approach is used to formulate the theoretical regulatory modes of the G-protein and ABA. We then apply a pattern matching approach to associate gene expression profiles with Boolean models. Our method allows us to address the following questions: (1) Are all of the classical modes of G-protein signaling observed and are any non-classical modes of G-protein signaling revealed? (2) Does the heterotrimeric G-protein have ABA-independent regulatory activity? (3) Is it possible that ABA and the G-protein co-regulate some processes, but ABA is not the signal activating the G-protein? (4) Is there system specificity of G-protein regulation of the transcriptome? (5) Does regulation at the level of the transcriptome recapitulate previously observed hypersensitivity or hyposensitivity of

developmental and dynamic transient responses to ABA in G-protein mutants?

# Results

According to the characteristics of the conditions under which we sample the transcriptome (G-protein subunit mutants with or without ABA), we use Boolean variables to denote the states of GPA1, AGB1, and ABA. Specifically, GPA1=0 and AGB1=0 represents the *agb1 gpa1* double mutant, GPA1=0 and AGB1=1 represents the *gpa1* mutant, GPA1=1 and AGB1=0 represents the *agb1* mutant, and GPA1=1 and AGB1=1 means wild type. ABA=1 indicates the presence of ABA (i.e. ABA treatment) and ABA=0 represents the absence of ABA (i.e. solvent control). The expression profile of a target gene in different genotypes or treatments is the result of the (indirect) regulatory activity of the G-protein and/or ABA, which can be seen as a truth vector of a Boolean function determined by the states of GPA1, AGB1, and/or ABA. Statistical analysis of gene expression data verifies the validity of this Boolean framework (Supplementary information 3). The regulatory modes of the G-protein are modeled by Boolean functions A(GPA1, AGB1), reflecting the effect of the G-protein on the expression level of a target gene (see Materials and methods). For example, the classical I mechanism of the G-protein can be represented by $A_2$(**GPA1, AGB1**)=**GPA1 and AGB1**. A target gene regulated by this G-protein mode is strongly expressed when both GPA1 and AGB1 are present (i.e. in the wild type) and weakly or not expressed in the mutant genotypes. The classical II mode of the G-protein corresponds to $A_5$(**GPA1, AGB1**)=**not GPA1 and AGB1**. A target gene regulated by this mode is strongly expressed in the *gpa1* mutant and shows weak or absent expression in other genotypes.

**Table I** Boolean regulatory modes of the G-protein and ABA and their corresponding gene expression patterns

| Co-regulatory modes of the G-protein and ABA | States of B(ABA, A(GPA1, AGB1)) db-$G\alpha$-$G\beta$-wt-<u>db-$G\alpha$-$G\beta$-wt</u> | Target gene expression db-$G\alpha$-$G\beta$-wt-<u>db-$G\alpha$-$G\beta$-wt</u> |
|---|---|---|
| $B_1$(ABA, A) = 0 | $[0, 0, 0, 0, 0, 0, 0, 0]$ | |
| $B_2$(ABA, A) = ABA and A | $[0, 0, 0, 0, A(0, 0), A(0, 1), A(1, 0), A(1, 1)]$ | |
| $B_3$(ABA, A) = ABA and not A | $[0, 0, 0, 0, \overline{A}(0,0), \overline{A}(0,1), \overline{A}(1,0), \overline{A}(1,1)]$ | |
| $B_4$(ABA, A) = ABA | $[0, 0, 0, 0, 1, 1, 1, 1]$ | |
| $B_5$(ABA, A) = not ABA and A | $[A(0,0), A(0,1), A(1,0), A(1,1), 0, 0, 0, 0]$ | |
| $B_6$(ABA, A) = A | $[A(0,0), A(0,1), A(1,0), A(1,1), A(0,0), A(0,1), A(1,0), A(1,1)]$ | |
| $B_7$(ABA, A) = (ABA or A) and not (ABA and A) | $[A(0,0), A(0,1), A(1,0), A(1,1), \overline{A}(0,0), \overline{A}(0,1), \overline{A}(1,0), \overline{A}(1,1)]$ | |
| $B_8$(ABA, A) = ABA or A | $[A(0,0), A(0,1), A(1,0), A(1,1), 1, 1, 1, 1]$ | |
| $B_{16}$(ABA, A) = 1 | $[1, 1, 1, 1, 1, 1, 1, 1]$ | |

$B_{17-i}$=not $B_i$, $i$=2, …, 8 are not shown because they are the logical negations of shown modes. The target gene expression in $B_2$, $B_3$, $B_5$, $B_6$, $B_7$, $B_8$ depends on the A(GPA1, AGB1) functions describing the G-protein regulation. For these functions, we show one example (different from case to case) of 14 possible target gene expression patterns. db-$G\alpha$-$G\beta$-wt denotes the order of genotypes, that is *agb1 gpa1* double mutant, *gpa1* mutant, *agb1* mutant and wild type, without ABA treatment. <u>db-$G\alpha$-$G\beta$-wt</u> denotes the corresponding order of genotypes with ABA treatment. $\overline{A}$(GPA1, AGB1) denotes the logical negation of A(GPA1, AGB1).

The co-regulation of a gene by the G-protein and ABA can be described as F(ABA, GPA1, AGB1)=$C_{ABA}$ + B(ABA, A(GPA1, AGB1)). $C_{ABA}$ is a constant that is non-zero if there is a G-protein-independent regulatory effect of ABA on a gene. B(ABA, A(GPA1, AGB1)), shown in Table I, is a nested Boolean function representing the possible combinatorial regulation of ABA and the G-protein. The complement of B(ABA, A(GPA1, AGB1)) functions determines a total of 72 non-trivial regulatory modes of the G-protein and/or ABA (see Materials and methods). Together with $C_{ABA}$, F(ABA, GPA1, AGB1) is able to identify 142 regulatory modes of the G-protein and ABA, which can be classified into five main regulatory categories (see Materials and methods): **ABA-only** regulation of transcript levels, **G-protein-only** regulation, G-protein–ABA **additive** regulation, ABA–G-protein **combinatorial** regulation, and ABA–G-protein **mixed** regulation. For example, **$B_2$(ABA, $A_2$(GPA1, AGB1))=ABA and (GPA1 and AGB1)** is one example of combinatorial regulation by the G-protein and ABA in which both G-protein subunits and ABA are required for high gene expression. **$B_4$(ABA, A(GPA1, AGB1))=ABA** exemplifies regulation by ABA independent of the status of the G-protein. **$C_{ABA}$ + $B_6$(ABA, $A_6$(GPA1, AGB1))=$C_{ABA}$ + AGB1** is one example of G-protein–ABA additive regulation, namely regulation by the G-protein β subunit with a separate additive input from ABA.

Each regulatory mode represented by A(GPA1, AGB1) or B(ABA, A(GPA1, AGB1)) corresponds to an idealized Boolean expression pattern. To avoid potential biases in binarizing expression levels, instead of correlating absolute expression profiles, our method is based on the differential expression patterns of genes. Figure 2 illustrates four examples of idealized expression patterns and idealized differential expression patterns of genes governed by Boolean functions. To associate genes with regulatory modes of the G-protein and/or ABA, we construct the real differential expression pattern of each gene and adopt a correlation-based pattern matching approach to examine whether the real differential

expression pattern of the gene is consistent with one of the idealized differential expression patterns given by the Boolean regulatory modes of the G-protein and/or ABA (see Materials and methods). By looking at the number of genes belonging to each regulatory mode, we are able to gauge the plausibility of each mode.

## Biologically plausible modes of G-protein regulation

Our first aim is to determine which of the theoretically possible G-protein regulatory modes, as reflected by the 14 Boolean functions $A_2$(GPA1, AGB1) to $A_{15}$(GPA1, AGB1), are experimentally supported by the transcriptome data. We separately consider four combinations of tissues and treatment, that is guard cells or leaves with or without ABA treatment. Using stringent thresholds for differential expression and correlation with idealized patterns (see Materials and methods), a total of 107 G-protein-regulated genes are identified in guard cells without ABA treatment, and 51 G-protein-regulated genes are identified in guard cells with ABA treatment. There are 103 G-protein-regulated genes in leaves without ABA treatment, and 143 G-protein-regulated genes in leaves with ABA treatment. These genes are listed in the Supplementary file Sup1.xls.

We observe a number of G-protein-regulated genes in the absence of ABA in both guard cells and leaves, indicating that the G-protein has regulatory activity independent of ABA treatment. Only the two knocked-out genes, *GPA1* and *AGB1*, are common to the G-protein-regulated gene sets in guard cells and leaves in either of the two ABA conditions, suggesting system specificity of G-protein action. The distribution of genes in each G-protein regulatory mode in guard cells and leaves, with and without ABA treatment, is given in Figure 3. Note that it is possible that the genes in $A_i$ and $A_{17-i}$ are regulated by the G-protein through the same signaling pathway, for example for $A_2$ and $A_{15}$, GPA1 and AGB1 could synergistically regulate a mediator (e.g. a transcription factor or other regulator of gene expression
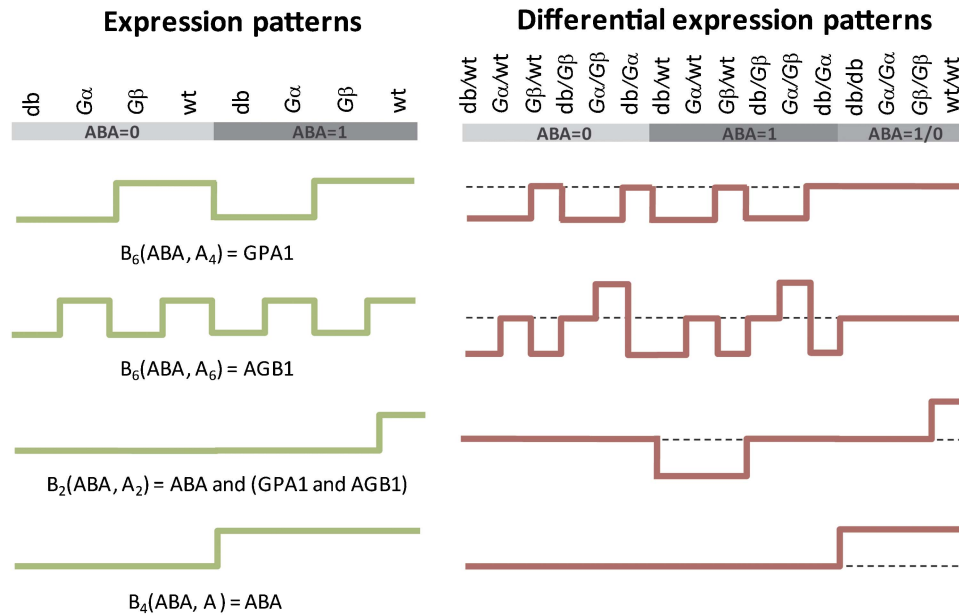
**Figure 2** Examples of idealized differential expression patterns determined by the regulatory modes of the G-protein and/or ABA. The lines on the left panel represent idealized gene expression patterns such as those in Table I. In the corresponding differential expression patterns in the right panel, the first six elements correspond to genotype comparisons in the absence of ABA, the second six elements correspond to genotype comparisons in the presence of ABA, and the last four elements correspond to comparison of ± ABA treatment within the same genotype. The dashed horizontal line corresponds to no differential expression. Segments below the dashed line indicate downregulation in the first condition versus the second condition and segments above it indicate upregulation.
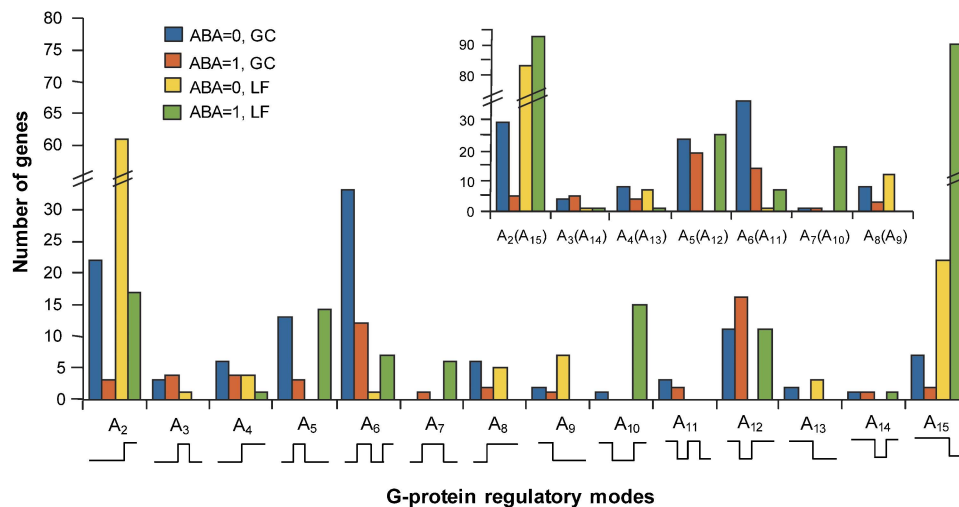


**Figure 3** Numbers of genes in each G-protein regulatory mode specified by an A(GPA1, AGB1) function in the four data sets. Blue and red bars correspond to the number of genes regulated in guard cells in the absence and presence of ABA, respectively. Yellow and green bars correspond to the number of genes regulated in leaves in the absence and presence of ABA, respectively. Idealized expression patterns for each mode are indicated below each A(GPA1, AGB1) function; each idealized expression pattern consists of four connected segments corresponding to genotypes; genotype order is as in Figure 2. The inset shows the number of genes in merged regulatory modes, each including an A(GPA1, AGB1) function and its logical negation.

such as a microRNA), which activates the genes in $A_2$ and inhibits the genes in $A_{15}$. To gain insight into the overall abundance of possible G-protein regulatory mechanisms, we plot the cumulative number of genes in each G-protein regulatory group by merging the genes present in $A_{17-i}$ with those in $A_i$, as shown in the inset of Figure 3 (see also Supplementary information 9 for the interpretation of the relationship between an A(GPA1, AGB1) function and its negation).

From Figure 3, we can see that $A_2$=**GPA1 and AGB1** (and/or $A_{15}$=**not $A_2$=not GPA1 or not AGB1**) is a popular G-protein regulatory mode in both guard cells and leaves. This regulatory mode requires the presence of both Gα and Gβγ subunits, and thus represents the classical Ia and Ib mechanisms, in which both Gα and Gβ(γ) are required. We note that, formally, this mode is also consistent with regulation by a non-dissociated heterotrimer (Klein *et al*, 2000; Frank *et al*, 2005; Digby *et al*, 2008). The regulatory

mechanism $A_5$=**not GPA1 and AGB1** (and/or $A_{12}$=**not $A_5$=GPA1 or not AGB1**) is also strongly supported in all conditions except in leaves without ABA treatment. This mode is consistent with the classical II G-protein regulatory mechanism, based on signaling by the freed Gβγ subunit, as it incorporates the fact that in the *gpa1* mutant, Gβγ will be available to regulate downstream genes. The G-protein regulatory mode $A_6$=**AGB1** (and/or $A_{11}$=**not AGB1**) is prevalent in guard cells but less so in leaves. This Boolean function represents a non-classical regulatory mechanism in which signaling by Gβ(γ) occurs, and this signaling is not regulated in any way by Gα; in other words, at no point during this signaling mechanism does Gβ(γ) assemble with Gα into a heterotrimer. The functions $A_3$=**GPA1 and not AGB1,** $A_{14}$=**not $A_3$=not GPA1 or AGB1** are not supported, suggesting that Gα does not regulate downstream effectors in the absence of Gβγ. The functions $A_4$=**GPA1**, $A_{13}$=**not GPA1** are also very weakly supported, suggesting that Gα signaling unaffected by the presence or absence of Gβγ is rare. The function $A_8$=**GPA1 or AGB1** (and/or $A_9$=**not $A_8$=not GPA1 and not AGB1**) is weakly supported in the absence of ABA, but its low abundance suggests that regulation of a target equally well by either free Gα or free Gβγ is relatively rare. Interestingly, the function $A_7$=**(GPA1 or AGB1) and not (GPA1 and AGB1)** (and/or $A_{10}$=**not $A_7$=(GPA1 and AGB1) or not (GPA1 or AGB1)**) is well represented in leaves with ABA treatment. The complex combinatorial relationship between Gα and Gβγ underlying this function suggests the involvement of more than one G-protein effector.

## Co-regulatory modes of the G-protein and ABA

We next combine the data with and without ABA treatment and determine which of the theoretically possible (co)regulatory modes of G-protein and ABA are experimentally supported by the

transcriptome data. By using stringent thresholds for association of genes with regulatory modes (see Materials and methods), 71 ABA–G-protein combinatorially regulated genes, 29 ABA–G-protein mixed regulated genes, 28 G-protein-only regulated genes, and 3 G-protein–ABA additively regulated genes are identified in guard cells. In contrast, in leaves 471 ABA–G-protein combinatorially regulated genes, 78 ABA–G-protein mixed regulated genes, and only 3 G-protein-only regulated genes are identified. We also obtain 467 ABA-only induced genes and 246 ABA-only repressed genes in guard cells, and 309 ABA-only induced genes and 94 ABA-only repressed genes in leaves. All of these genes are listed in the Supplementary files Sup2.xls for guard cells and Sup3.xls for leaves.
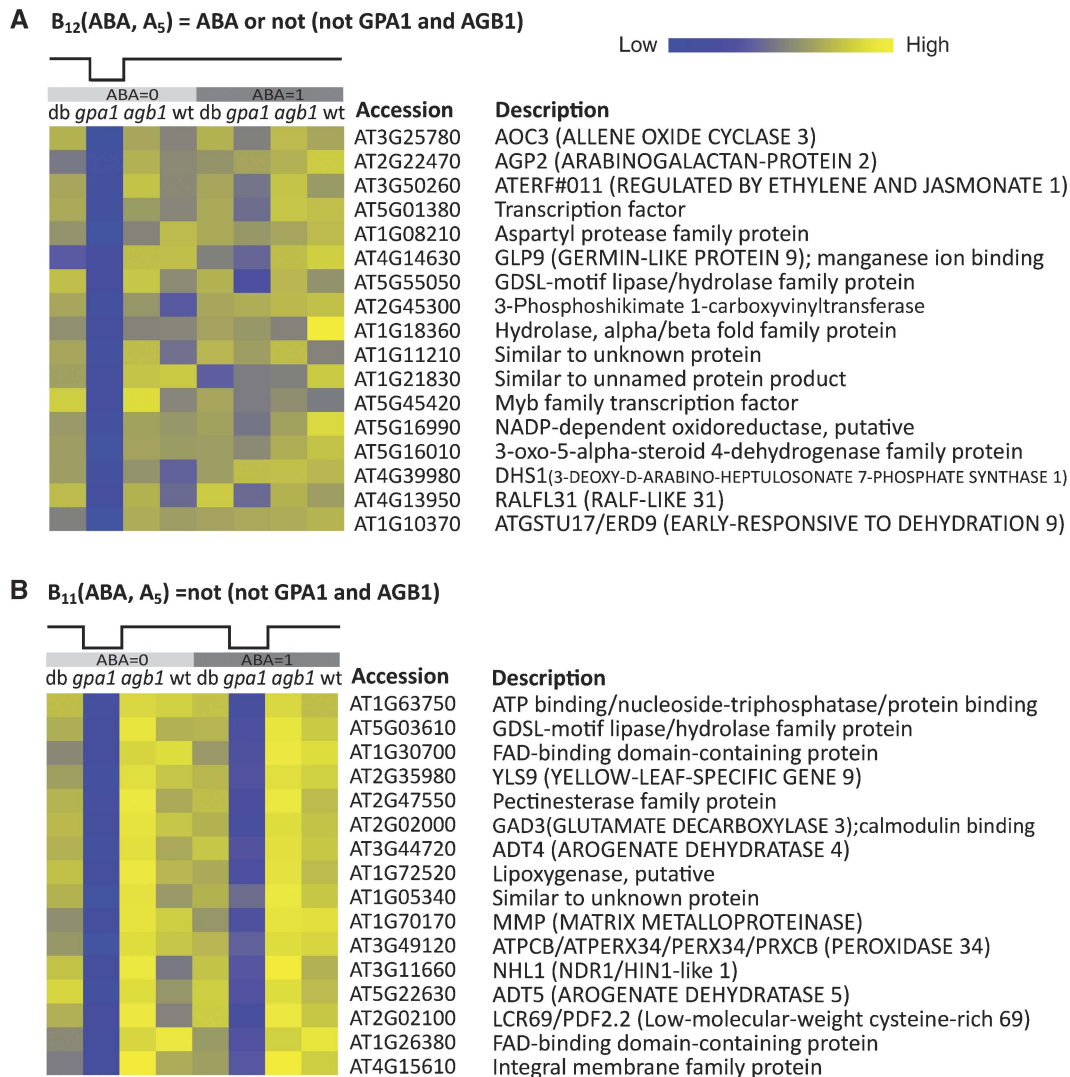
To focus on the combinatorial G-protein–ABA regulatory modes encapsulated in the B(ABA, A(GPA1, AGB1)) function, we cumulate ABA–G-protein combinatorially regulated genes and ABA–G-protein mixed regulated genes that are governed by the same Boolean function. Similarly, the genes governed by the same Boolean function in G-protein-only regulation and G-protein–ABA additive regulation are also merged. We find that 33 out of the 72 theoretically possible regulatory modes are unsupported (have 0 genes) in guard cells and 34 regulatory modes are unsupported in leaves. The distribution of genes in each supported regulatory mode of the G-protein and ABA in guard cells and leaves is illustrated in Supplementary Figures S10–S12. In Table II, we summarize the regulatory modes that govern >10% of the genes in the relevant category (ABA–G-protein combinatorial/mixed regulation, or G-protein-only/G-protein–ABA additive regulation).

We observe that in guard cells, $A_5$ (and/or $A_{12}$)-related regulatory modes are the most representative: $B_{12}$(**ABA, $A_5$**)=**ABA or not (not GPA1 and AGB1)**=**ABA or not $A_5$**, $B_{15}$(**ABA, $A_5$**)=**not ABA or not $A_5$**, and $B_5$(**ABA, $A_5$**)=**not ABA and $A_5$**. These co-regulatory modes of the G-protein and ABA are consistent with Figure 1D where the mediator M1 is

**Table II** Well-supported G-protein regulatory modes in guard cells and leaves

| Boolean function | Regulatory mode | Target gene expression | No. of genes | Tissue |
|---|---|---|---|---|
| $B_{15}$(ABA, $A_5$) = not ABA or not (not GPA1 and AGB1) | Classical II | | 11 (11%) | GC |
| $B_5$(ABA, $A_5$) = not ABA and (not GPA1 and AGB1) | Classical II | | 17 (17%) | GC |
| $B_{12}$(ABA, $A_5$) = ABA or not (not GPA1 and AGB1) | Classical II | | 17 (17%) | GC |
| $B_{11}$(ABA, $A_5$) = not (not GPA1 and AGB1) | Classical II | | 16 (52%) | GC |
| $B_6$(ABA, $A_6$) = AGB1 | Non-classical | | 7 (23%) | GC |
| $B_2$(ABA, $A_2$) = ABA and (GPA1 and AGB1) | Classical I | | 86 (16%) | LF |
| $B_{15}$(ABA, $A_2$) = not ABA or not (GPA1 and AGB1) | Classical I | | 154 (46%) | LF |
| $B_5$(ABA, $A_2$) = not ABA and (GPA1 and AGB1) | Classical I | | 189 (34%) | LF |
| $B_{12}$(ABA, $A_2$) = ABA or not (GPA1 and AGB1) | Classical I | | 58 (11%) | LF |

Well-supported modes are defined as those containing ⩾10% of the genes within one of two regulatory categories: G-protein-only/G-protein–ABA additive regulation, and ABA–G-protein combinatorial/mixed regulation. The third column shows the idealized expression pattern corresponding to the mode. The pattern is given in the same genotype order as in Table I and Figure 2; dashed lines indicate the pattern in the absence of ABA and solid lines indicate the pattern in the presence of ABA. The fourth column indicates the number of genes associated with the regulatory mode and their percentage within the relevant category; the fifth column indicates the relevant tissue (GC=guard cells; LF=leaves).

**A** $B_{12}(ABA, A_5) = ABA$ or not (not GPA1 and AGB1)

Low ▮▮▮▮▮ High



| | Accession | Description |
|---|---|---|
| | AT3G25780 | AOC3 (ALLENE OXIDE CYCLASE 3) |
| | AT2G22470 | AGP2 (ARABINOGALACTAN-PROTEIN 2) |
| | AT3G50260 | ATERF#011 (REGULATED BY ETHYLENE AND JASMONATE 1) |
| | AT5G01380 | Transcription factor |
| | AT1G08210 | Aspartyl protease family protein |
| | AT4G14630 | GLP9 (GERMIN-LIKE PROTEIN 9); manganese ion binding |
| | AT5G55050 | GDSL-motif lipase/hydrolase family protein |
| | AT2G45300 | 3-Phosphoshikimate 1-carboxyvinyltransferase |
| | AT1G18360 | Hydrolase, alpha/beta fold family protein |
| | AT1G11210 | Similar to unknown protein |
| | AT1G21830 | Similar to unnamed protein product |
| | AT5G45420 | Myb family transcription factor |
| | AT5G16990 | NADP-dependent oxidoreductase, putative |
| | AT5G16010 | 3-oxo-5-alpha-steroid 4-dehydrogenase family protein |
| | AT4G39980 | DHS1 (3-DEOXY-D-ARABINO-HEPTULOSONATE 7-PHOSPHATE SYNTHASE 1) |
| | AT4G13950 | RALFL31 (RALF-LIKE 31) |
| | AT1G10370 | ATGSTU17/ERD9 (EARLY-RESPONSIVE TO DEHYDRATION 9) |

**B** $B_{11}(ABA, A_5) =$ not (not GPA1 and AGB1)



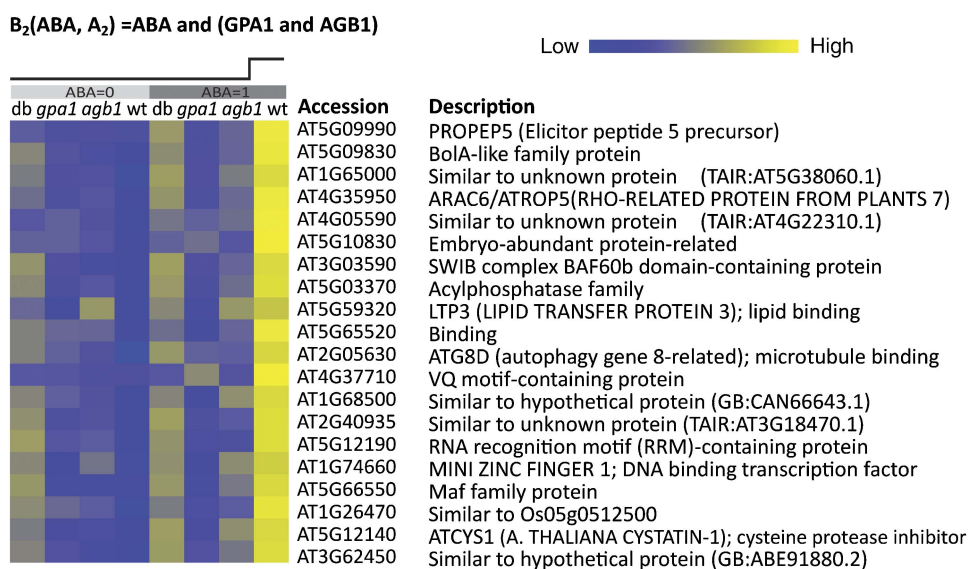| | Accession | Description |
|---|---|---|
| | AT1G63750 | ATP binding/nucleoside-triphosphatase/protein binding |
| | AT5G03610 | GDSL-motif lipase/hydrolase family protein |
| | AT1G30700 | FAD-binding domain-containing protein |
| | AT2G35980 | YLS9 (YELLOW-LEAF-SPECIFIC GENE 9) |
| | AT2G47550 | Pectinesterase family protein |
| | AT2G02000 | GAD3 (GLUTAMATE DECARBOXYLASE 3); calmodulin binding |
| | AT3G44720 | ADT4 (AROGENATE DEHYDRATASE 4) |
| | AT1G72520 | Lipoxygenase, putative |
| | AT1G05340 | Similar to unknown protein |
| | AT1G70170 | MMP (MATRIX METALLOPROTEINASE) |
| | AT3G49120 | ATPCB/ATPERX34/PERX34/PRXCB (PEROXIDASE 34) |
| | AT3G11660 | NHL1 (NDR1/HIN1-like 1) |
| | AT5G22630 | ADT5 (AROGENATE DEHYDRATASE 5) |
| | AT2G02100 | LCR69/PDF2.2 (Low-molecular-weight cysteine-rich 69) |
| | AT1G26380 | FAD-binding domain-containing protein |
| | AT4G15610 | Integral membrane family protein |

**Figure 4** Gene expression patterns illustrating the classical II G-protein regulatory mode coupled with regulation by ABA (**A**), or independent of ABA (**B**) in guard cells. The heat map is generated with the matrix2png software (Pavlidis and Noble, 2003). The rows of each pattern correspond to genes, and the columns correspond to conditions, that is four genotypes, *agb1 gpa1* double mutant (db), *gpa1* mutant, *agb1* mutant, wild type (wt), without or with ABA treatment. The header of each group of patterns indicates the Boolean rule and the idealized expression pattern associated with the group. The expression level for each gene is extracted from the original expression profiles by subtracting $C_{ABA}$ (if relevant), then averaging over the three replicates and normalizing across the samples such that the mean is 0 and the s.d. is 1. TAIR (http://www.arabidopsis.org/) derived accession numbers and a brief description of the genes are shown next to the heat map.

regulated by the classical II G-protein regulatory mechanism dependent on the Gβγ subunit on its release from Gα, and the mediator M2 is combinatorially regulated by ABA and M1. Figure 4A uses the commonly used 'heat map' portrayal to illustrate one of these well-supported classical II modes, $B_{12}(ABA, A_5)$, in which the absence of Gα frees Gβγ to downregulate downstream genes, in this instance with the effect of Gα knockout on gene expression dependent on the absence or presence of ABA. In guard cells, G-protein-only signaling and G-protein–ABA additive regulation, where the relative effect of subunit knockout is the same regardless of the presence or absence of ABA, are also possible. The classical II mode $B_{11}(ABA, A_5)=$ **not (not GPA1 and AGB1)** (Figure 4B) and the non-classical mode $B_6(ABA, A_6)=$**AGB1** (Figure 5) are the two most supported G-protein-only/G-protein–ABA additive regulatory modes.

In leaves, $A_2$ (and/or $A_{15}$)-related regulatory modes are the most prevalent: $B_2(ABA, A_2)=$**ABA and (GPA1 and AGB1)**= **ABA and** $A_2$, $B_{15}(ABA, A_2)=$**not ABA or not** $A_2$, $B_5(ABA, A_2)=$**not ABA and** $A_2$, and $B_{12}(ABA, A_2)=$**ABA or not** $A_2$. These regulatory modes are consistent with Figure 1D where the mediator M1 is synergistically regulated by GPA1 and AGB1 (classical I G-protein regulatory mechanisms) and the second mediator M2 is combinatorially regulated by M1 and ABA. Figure 6 illustrates one of these well-supported modes, $B_2(ABA, A_2)=$**ABA and (GPA1 and AGB1)**, in which gene expression is highest in wild type plus ABA, due to a requirement for both a wild-type G-protein complement and ABA treatment to elevate transcript levels. There are only three genes in the G-protein-only category in leaves, and two of them are the G-protein subunits *GPA1* and *AGB1* included by default; the third is At3G12730, a MYB family transcription factor. This

$B_6(ABA, A_6) = AGB1$



| Accession | Description |
|-----------|-------------|
| AT4G34460 | AGB1 (GTP BINDING PROTEIN BETA 1); nucleotide binding |
| AT5G24110 | WRKY30 (WRKY DNA-binding protein 30) |
| AT4G02410 | Lectin protein kinase family protein |
| AT1G20310 | Similar to unknown protein |
| AT3G17700 | CNBT1; calmodulin binding/ion channel |
| AT5G42010 | WD-40 repeat family protein |
| AT2G40180 | ATHPP2C5; protein serine/threonine phosphatase |

**Figure 5** Gene expression pattern for a novel G-protein regulatory mode, AGB1, in guard cells. TAIR (http://www.arabidopsis.org/) derived accession numbers and a brief description of the genes are shown next to the heat map.

$B_2(ABA, A_2) = ABA$ and (GPA1 and AGB1)



| Accession | Description |
|-----------|-------------|
| AT5G09990 | PROPEP5 (Elicitor peptide 5 precursor) |
| AT5G09830 | BolA-like family protein |
| AT1G65000 | Similar to unknown protein    (TAIR:AT5G38060.1) |
| AT4G35950 | ARAC6/ATROP5(RHO-RELATED PROTEIN FROM PLANTS 7) |
| AT4G05590 | Similar to unknown protein    (TAIR:AT4G22310.1) |
| AT5G10830 | Embryo-abundant protein-related |
| AT3G03590 | SWIB complex BAF60b domain-containing protein |
| AT5G03370 | Acylphosphatase family |
| AT5G59320 | LTP3 (LIPID TRANSFER PROTEIN 3); lipid binding |
| AT5G65520 | Binding |
| AT2G05630 | ATG8D (autophagy gene 8-related); microtubule binding |
| AT4G37710 | VQ motif-containing protein |
| AT1G68500 | Similar to hypothetical protein (GB:CAN66643.1) |
| AT2G40935 | Similar to unknown protein (TAIR:AT3G18470.1) |
| AT5G12190 | RNA recognition motif (RRM)-containing protein |
| AT1G74660 | MINI ZINC FINGER 1; DNA binding transcription factor |
| AT5G66550 | Maf family protein |
| AT1G26470 | Similar to Os05g0512500 |
| AT5G12140 | ATCYS1 (A. THALIANA CYSTATIN-1); cysteine protease inhibitor |
| AT3G62450 | Similar to hypothetical protein (GB:ABE91880.2) |

**Figure 6** Gene expression pattern for a classical I G-protein regulatory mode coupling with ABA in leaves. The 20 genes with the highest correlation scores (out of 86) are shown. TAIR (http://www.arabidopsis.org/) derived accession numbers and a brief description of the genes are shown next to the heat map.

suggests that in leaves there are very few genes that are regulated by the G-protein independently of ABA signaling and instead the co-regulatory mode of the G-protein and ABA is prevalent.

To test the robustness of the above conclusions, we repeated the analysis with a looser genotype threshold and found no new representative gene groups, only more genes for the originally representative regulatory modes (see Supplementary Figures S10–S12). These results indicate that these regulatory modes are biologically plausible and there indeed exists a coupling between transcription regulation by the G-protein and ABA in both guard cells and leaves.

## Relating co-regulated genes to the G-protein and ABA signaling pathways

The regulatory modes and ABA additive effects identified in our analysis also enable determination of the most plausible G-protein and ABA signaling pathways (see Figure 1; Supplementary Figure S1). ABA-only regulated genes

(i.e. $B_4(ABA, A_i) = ABA$, $B_{13}(ABA, A_i) = $ **not** $ABA$, $i = 2, \ldots, 15$), correspond to the pathway in Figure 1A, indicating that there is a G-protein-independent effect of ABA, and the status of the G-protein is irrelevant. This regulatory mode is strongly supported in both guard cells and leaves: it describes 713 genes in guard cells (467 upregulated and 246 downregulated, see Sheet 5 of the Supplementary file Sup2.xls) and 403 genes in leaves (309 upregulated and 94 downregulated, see Sheet 5 of the Supplementary file Sup3.xls). This result indicates that there are G-protein-independent ABA signaling pathways in both guard cells and leaves. The signaling pathway of Figure 1B means that the G-protein is turned on by an ABA-independent signal, and there is no G-protein-independent effect of ABA on gene expression. This regulatory mode corresponds to G-protein-only regulated genes (i.e. $B_6(ABA, A_i) = A_i$, $B_{11}(ABA, A_i) = $ **not** $A_i$, $i = 2, \ldots, 15$ with non-significant $C_{ABA}$), and is exhibited by 28 genes in guard cells (see Sheet 3 of Sup2.xls) but only three genes (of which two are *GPA1* and *AGB1*) in leaves (see Sheet 3 of Sup3.xls). These results corroborate the existence of an ABA-independent regulatory activity of the G-protein in

guard cells, through both classical ($B_{11}$(ABA, $A_5$)) and non-classical ($B_6$(ABA, $A_6$)) regulatory modes. Interestingly, these ABA-independent G-protein regulatory modes are not supported in leaves.

The signaling pathway of Figure 1C means that in addition to ABA-independent G-protein regulation, there is also a G-protein-independent effect of ABA. This regulatory mode corresponds to G-protein–ABA additive regulated genes (i.e. $B_6$(ABA, $A_i$) + $C_{ABA}$ = $A_i$ + $C_{ABA}$, $B_{11}$(ABA, $A_i$) + $C_{ABA}$ = not $A_i$ + $C_{ABA}$) and represents an additive cross-talk, converging on the target gene, between ABA and the signal that activates the G-protein. This regulatory mode has three genes in guard cells (see Sheet 4 of Sup2.xls) and no genes in leaves, indicating that additive regulation by G-protein signaling plus G-protein-independent ABA signaling is rare.

The signaling pathway shown in Figure 1D represents the case where the G-protein is turned on by an ABA-independent signal, and there is a combinatorial G-protein–ABA effect. This pathway corresponds to ABA–G-protein combinatorially regulated genes with non-significant $C_{ABA}$ (i.e. $B_2$(ABA, $A_i$) = $A_i$ and ABA, $B_8$(ABA, $A_i$) = $A_i$ or ABA, $B_5$(ABA, $A_i$) = $A_i$ and not ABA, and $B_{14}$(ABA, $A_i$) = $A_i$ or not ABA, $i$ = 2,...,15) and represents a combinatorial cross-talk that converges on the mediator M2 (see Figure 1D). This regulatory mode is observed for 71 genes in guard cells (see Sheet 1 of Sup2.xls) and 471 genes in leaves (see Sheet 1 of Sup3.xls). A similar pathway (Supplementary Figure S1E) is that in addition to the combinatorial G-protein–ABA regulatory effect, there is also a G-protein-independent effect of ABA (i.e. ($A_i$ and ABA) + $C_{ABA}$, ($A_i$ or ABA) + $C_{ABA}$, ($A_i$ and not ABA) + $C_{ABA}$, and ($A_i$ or not ABA) + $C_{ABA}$, $i$ = 2,...,15). This regulatory mode corresponds to ABA–G-protein mixed regulated genes and is supported by 29 genes in guard cells (see Sheet 2 of Sup2.xls) and 78 genes in leaves (see Sheet 1 of Sup3.xls), indicating that cross-talk between ABA–G-protein signaling and G-protein-independent ABA signaling occurs.

In mammalian systems, it is well established that a given Gα can interact with more than one specific GPCR and thus can be regulated by more than one ligand (Leaney and Tinker, 2000). In *Arabidopsis*, if the signal that catalyzes the dissociation of Gα and Gβγ and activates the G-protein were ABA (see Supplementary Figure S1F–I), then without ABA treatment, the G-protein would be off and would have no regulatory activity (described by $B_2$(ABA, $A_i$) = $A_i$ and ABA, $B_{14}$(ABA, $A_i$) = $A_i$ or not ABA, $i$ = 2,...,15). In addition to such regulation of the G-protein by ABA, there could be a G-protein-independent effect of ABA, or/and a G-protein–ABA combinatorial regulatory effect. The genes regulated by these modes would form an unknown subset of ABA–G-protein combinatorially or mixed regulated genes, that is all or a fraction of the genes that have no differential expression with respect to genotypes in the absence of ABA (see Supplementary information 11 for a detailed description). The maximum possible number of genes regulated by a mode in which ABA is the signal activating the G-protein is 38 in guard cells and 287 in leaves. In other words, these genes are consistent with the possibility of ABA serving as a ligand for a GPCR. On the other hand, for the specific case where the G-protein functions according to the classical mechanism II, we can conclude that our analysis does not validate the scenario in which ABA is the signal that activates the G-protein (see Supplementary information 11 for the reasoning behind this statement). Finally, the 62 genes in guard cells and 262 genes in leaves consistent with ABA–G-protein combinatorial or mixed regulation but not consistent with ABA activation of the G-protein are a clear indication that the G-protein can be activated by a signal other than ABA.

## ABA hypo-/hypersensitivity of gene regulation in G-protein mutants

For guard cell responses published in the literature to date, namely for ABA inhibition of stomatal opening and inward K$^+$ channel regulation, *gpa1* and *agb1* mutants are equally hyposensitive to ABA (i.e. less sensitive than wild type) (Wang *et al*, 2001; Coursol *et al*, 2003; Fan *et al*, 2008). By contrast, in seed germination and post-germination seedling growth, G-protein subunit mutants are un-equally hypersensitive to ABA (i.e. more sensitive than wild type): the *agb1* single and the *agb1 gpa1* double mutants are more hypersensitive than the *gpa1* single mutant (Pandey *et al*, 2006). There are no extant studies regarding mature leaf sensitivity to ABA in the G-protein mutants versus wild type, nor any transcriptome-wide studies on the sensitivities of gene regulation to ABA in the G-protein mutants in any cell or tissue type. Here, we use our transcriptome data to assess and compare gene regulation by ABA in guard cells and leaves of the G-protein mutants.

ABA-only regulated genes by definition have equal responses to ABA in all mutant genotypes. G-protein-only regulated genes and G-protein–ABA additively regulated genes by definition have no or equal response to ABA in all mutant genotypes; thus, these genes are not useful for revealing ABA hyposensitivity or hypersensitivity. Therefore, we investigate those genes that exhibit ABA–G-protein co-regulation, comparing each ABA–G-protein-regulated gene's expression change in response to ABA in the wild-type genotype versus each mutant genotype. We define hyposensitivity to ABA as the case when the gene has lesser up/downregulation in response to ABA in a mutant as compared with wild type. Similarly, we define hypersensitivity to ABA as the case when the gene has greater up/downregulation in response to ABA in the mutant as compared with wild type. We also determine the numbers of genes that have similar responses in combinations of two mutant genotypes, or all three of them.

The fraction of regulated genes exhibiting hyposensitive or hypersensitive response to ABA in guard cells and leaves of the mutants are listed in Table III. To determine the significance of these results, we compare these fractions to those generated using a random background (given in Supplementary Table S3), in which the genes are randomly scattered over all Boolean rules that reflect ABA–G-protein co-regulation. Significance values are computed by a binomial cumulative distribution. We find that in guard cells, equal hyposensitivity in all mutants (the last row) and hypersensitivity in the *gpa1* mutant are significant ($P$ = 5 × 10$^{-5}$ and $P$ = 3 × 10$^{-8}$, respectively). In leaves, ABA hyposensitivity of gene expression in the three individual mutants (the first three rows) and equal

**Table III** Hyposensitivity and hypersensitivity of gene expression in response to ABA in the mutants

| Mutants | Hyposensitivity | | Hypersensitivity | | Wild-type response | |
|---|---|---|---|---|---|---|
| | GC (%) | LF (%) | GC (%) | LF (%) | GC (%) | LF (%) |
| *agb1* mutant | 26.7 | **87.0** | 8.6 | 4.5 | **64.8** | 8.5 |
| *gpa1* mutant | 21.0 | **90.3** | **53.3** | 7.1 | 25.7 | 2.6 |
| double mutant | 26.7 | **87.5** | 21.9 | 4.3 | **51.4** | 8.1 |
| *agb1* and *gpa1* mutants | 19.1 | **86.5** | 1.9 | 3.8 | 11.4 | 1.4 |
| *agb1* and double mutants | 24.8 | **84.8** | 7.6 | 3.3 | **48.6** | 4.7 |
| *gpa1* and double mutants | 19.1 | **86.3** | 5.7 | 3.1 | 1.9 | 0.2 |
| All three mutants | **18.1** | **84.3** | 1.9 | 2.8 | — | — |

The percentage of ABA–G-protein co-regulated genes that exhibit ABA hyper-/hyposensitivity or wild-type response is given in each mutant or mutant combinations. The percentages that are statistically significant as compared with a random background ($P < 10^{-4}$) are shown in boldface. GC=guard cells, LF=leaves.

**Table IV** Functional enrichment in G-protein regulatory modes

| Regulatory modes | Expression patterns (ABA=0, ABA=1) | Total No. | Enriched functions | No. | *P*-value |
|---|---|---|---|---|---|
| **GC** | | | | | |
| $A_6$(GPA1, AGB1) = AGB1 | | 33 | Phosphate metabolism | 12 | 6E-7 |
| | | | Modification by phosphorylation | 11 | 9E-8 |
| $A_{11}$(GPA1, AGB1) = not AGB1 | | 2 | Plant defense response | 2 | 2E-5 |
| $B_8$(ABA, $A_4$) = ABA or GPA1 | | 2 | Extracellular/secretion proteins | 2 | 1E-5 |
| $B_{12}$(ABA, $A_5$) = ABA or not ( not GPA1 and AGB1) | | 17 | Aromatic compound metabolism | 3 | 5E-6 |
| | | | Metabolism of dehydroquinic acid | 2 | 6E-5 |
| $B_{11}$(ABA, $A_5$) = not (not GPA1 and AGB1) | | 16 | Biosynthesis of phenylalanine | 2 | 1E-5 |
| | | | Disease, virulence and defense | 6 | 1E-7 |
| **LF** | | | | | |
| $A_6$(GPA1, AGB1) = AGB1 | | 7 | Response to biotic stimulus | 3 | 4E-5 |
| $B_5$(ABA, $A_2$) = not ABA and (GPA1 and AGB1) | | 189 | Nucleotide/nucleoside binding | 33 | 1E-8 |
| | | | Chloroplast | 52 | 7E-9 |
| $B_{12}$(ABA, $A_2$) = ABA or not (GPA1 and AGB1) | | 58 | Calcium binding | 5 | 3E-5 |

The third column indicates the total number of genes corresponding to the regulatory mode indicated in the first column and the fifth column indicates how many of these genes have the annotation given in the fourth column. GC=guard cells, LF=leaves.

hyposensitivity in all mutants (the last row) are highly significant ($P < 10^{-50}$ for all these cases). Using a less stringent differential expression threshold ($P_{geno} < 0.05$) leads to very similar results (data not shown).

## Functional preferences in G-protein regulatory modes

To examine whether specific G-protein regulatory modes tend to be involved in specific aspects of plant physiology, we perform functional analysis on the co-regulated genes in each G-protein regulatory mode using the FunCat functional catalogue of the MIPS database (Mewes *et al*, 2006). The *P*-values are calculated by the hypergeometric cumulative distribution and Bonferroni correction was applied for multi-

ple testing. We applied a highly stringent cutoff of $P < 10^{-4}$ and identified 12 functional categories significantly enriched in G-protein regulatory modes (acknowledging the caveat that absolute gene numbers are small in some categories), shown in Table IV. The genes in these significant functional categories are listed in Supplementary Tables S4 and S5 and in Supplementary information 13.

## Discussion

As we have described, there are multiple theoretical signaling pathways and regulatory modes involving the G-protein and/ or ABA. Through our identification of the Boolean rules governing genes (co)-regulated by the G-protein and ABA, we can relate these genes to 72 theoretical regulatory

modes (Table I) and the proposed nine signaling pathways (Supplementary Figure S1), providing insights into the biological plausibility of G-protein and ABA regulation and allowing us to address the questions posed in the Introduction.

## Are all of the classical modes of G-protein signaling supported and are any non-classical modes of G-protein signaling revealed?

Phenotypic analysis of Arabidopsis G-protein mutants has indicated that classic mammalian G-protein regulatory paradigms are also used during plant G-protein signaling. Our analysis strongly supports the conclusion that such classical paradigms also function in the regulation of the plant transcriptome. Our analysis indicates that most of the 70 theoretical regulatory mechanisms involving the G-protein (i.e. excluding two ABA-only regulatory modes) are unsupported and thus may be biologically untenable, at least in the biological systems investigated here. For example, 33 out of these 70 theoretical regulatory modes are unsupported (have 0 genes) and only five are strongly supported in guard cells; the corresponding numbers for leaves are 34 and 4 (Supplementary Figure S10–S12; Table II). These results lend credence to our analysis method, as it would be unexpected if G-protein regulation of the transcriptome were to proceed through entirely different mechanisms than the well-established mechanisms revealed by physiological analyses.

Notably, our analysis also reveals a novel G-protein regulatory mechanism: positive regulation by the $G\beta(\gamma)$ subunit independent of the presence or absence of the $G\alpha$ subunit is well supported in the guard cell transcriptome. According to this rule, **AGB1**, *agb1* plants exhibit a mutant phenotype, whereas *gpa1* plants are wild type for the same phenotype. Interestingly, among the numerous physiological and developmental responses that are G-protein regulated in Arabidopsis as judged by mutant analysis (Perfus-Barbeoch *et al*, 2004), there are a few phenotypes consistent with this pattern: *agb1* mutants are resistant to the protein glycosylation inhibitor, tunicamycin, whereas *gpa1* plants exhibit wild-type sensitivities (Wang *et al*, 2007). This response has been associated with AGB1 resident in the endoplasmic reticulum (Wang *et al*, 2007), suggesting that subcellular localization may be one mechanism through which $G\beta(\gamma)$ subunits are partitioned into heterotrimer-dependent versus heterotrimer-independent functions. Another plausible **AGB1** phenotype is altered root waving and skewing, which is observed in *agb1* mutants and not in *gpa1* mutants (Pandey *et al*, 2008).

To our knowledge, $G\beta(\gamma)$ function completely independent of any input from $G\alpha$ (i.e. $G\beta(\gamma)$ never assembles into a heterotrimer, even in the absence of agonist) has not been previously implicated in mammalian systems, except possibly for non-$G\gamma$-related signaling modes of the unusual $G\beta_5$ subunit (Dupre *et al*, 2009). This may be because the mechanism is unique to plants or it may be because the mechanism only occurs in specialized mammalian cell types or systems and thus has escaped detection. Alternatively, it may be that the complexity and redundancy of the G-protein complement in mammals has simply precluded discovery of this mechanism to date, whereas our use of a simpler model system

coupled with the ability to monitor thousands of phenotypes simultaneously enabled detection. Thus, future investigation of this mechanism in both plants and metazoans is warranted.

Conversely, our analysis indicates that the regulation by $G\alpha$ independent of the presence or absence of the $G\beta\gamma$ subunit is not supported by the microarray data. It has been proposed that GPA1, owing to fast kinetics of GDP/GTP exchange and a slow rate of GTP hydrolysis, would persist constitutively in the GTP-bound form, and thus could function independently from $G\beta(\gamma)$ (Johnston *et al*, 2007). In our own biochemical analyses (Pandey *et al*, 2009), we also observed slow rates of GTP hydrolysis by GPA1 but rates of GTP$\gamma$S binding were comparable to those for bovine $G\alpha$ assayed by the same method. Our analysis here does not indicate independent $G\alpha$ function, at least with regard to transcriptome regulation in guard cells and leaves.

## Does the heterotrimeric G-protein have ABA-independent regulatory activity? Is it possible that ABA and the G-protein co-regulate some processes, but ABA is not the signal activating the G-protein?

Previous physiological analyses of guard cell function have focused almost exclusively on the roles of heterotrimeric G-proteins in ABA signaling (Wang *et al*, 2001; Fan *et al*, 2008). Our transcriptome analysis now reveals that in this cell type, the heterotrimeric G-protein also has clear ABA-independent activity. Thus, when ABA–G-protein co-regulatory modes are evaluated, two G-protein-**only** functions are seen in guard cells: **GPA1 or not AGB1** (classical II) and **AGB1**. Guard cells actually sense a wide variety of signals, including light, $CO_2$, humidity, several hormones, and pathogens (Roelfsema and Hedrich, 2005; Israelsson *et al*, 2006; Shimazaki *et al*, 2007; Melotto *et al*, 2008; Acharya and Assmann, 2009). Our results now suggest that potential roles of G-proteins in these other signaling processes should also be evaluated. Indeed, there is already some evidence for heterotrimeric G-protein involvement in guard cell response to pathogens (Zhang *et al*, 2008b), as well as in guard cell development (Zhang *et al*, 2008a).

In guard cells, there is also evidence that ABA and the G-protein co-regulate sets of genes, consistent with known co-regulation by ABA and the G-protein at the physiological level. For the 62 genes corresponding to the functions $B_8(ABA, A_i) = A_i$ or $ABA$, $B_5(ABA, A_i) = A_i$ and not $ABA$ with or without a significant $C_{ABA}$, ABA can be excluded as the activating signal for the G-protein. An additional 38 guard cell genes exhibit regulation consistent with ABA being the signal that activates the G-protein; however, in the absence of additional biological information, these patterns are also consistent with ABA regulating signal transduction downstream of the G-protein: our current data do not allow us to distinguish between these two possibilities.

Our analysis does not support G-protein functions independent of ABA (G-protein-**only** regulatory modes) in mature leaves. There are no genes in the leaf transcriptome that exhibit an identical pattern of expression relative to genotype in both the presence and the absence of ABA. There is strong support, however, for both **mixed** and **combinatorial** regula-

**Table V** Representation of the classical I and classical II G-protein mechanisms in ABA–G-protein co-regulatory modes

| | GC | LF |
|---|---|---|
| *Classical I* | 21/100 | 503/549 |
| $B_2(ABA, A_2)$, $B_2(ABA, A_{15})$, $B_{14}(ABA, A_2)$, $B_{14}(ABA, A_{15})$, $B_5(ABA, A_2)$, $B_5(ABA, A_{15})$, $B_8(ABA, A_2)$, $B_8(ABA, A_{15})$ | | |
| *Classical II* | 51/100 | 27/549 |
| $B_2(ABA, A_5)$, $B_2(ABA, A_{12})$, $B_{14}(ABA, A_5)$, $B_{14}(ABA, A_{12})$, $B_5(ABA, A_5)$, $B_5(ABA, A_{12})$, $B_8(ABA, A_5)$, $B_8(ABA, A_{12})$ | | |

The first column indicates the B(ABA, A) functions related to the two classical G-protein mechanisms. In the second and third columns, the denominator of each ratio indicates the total number of genes assigned to the G-protein–ABA co-regulatory modes in the given tissue type and the numerator indicates the number of genes supporting the given classical mechanism. GC=guard cells, LF=leaves.

tion by ABA and the G-protein. To date, ABA responses of rosette leaves have not been studied in the G-protein mutants, and our results strongly suggest that such studies would be useful for deciphering novel G-protein/ABA signaling pathways.

## Is there system specificity of G-protein regulatory mechanisms?

To account for the diversity of physiological and developmental phenotypes observed in knockout mutants of plant G-proteins, despite the small numbers of G-protein subunits encoded in plant genomes, it has been proposed that there is specificity in how G-protein signaling components are used in different cells and tissues. Our analyses corroborate this hypothesis, through several lines of reasoning. First, the regulatory modes that are strongly supported by our analysis differ in guard cells versus leaves. This is true both when we evaluate G-protein regulatory modes without considering ABA effects (Figure 3) and when we combine all the data together to analyze G-protein/ABA co-regulation (Table II). For example, in both the former and the latter analyses, **AGB1** is an observed regulatory mode in guard cells but not in leaves and, in fact, for leaves there is no G-protein-**only** regulatory mode that is supported when G-protein/ABA co-regulation is considered. These results suggest that heterotrimeric G-proteins are actually more involved in ABA signaling in leaves than in guard cells, an interpretation that is consistent with the stronger ABA hyposensitivity exhibited by the leaf transcriptome than the guard cell transcriptome of the G-protein mutants (see next section). This new finding was unexpected, given that ABA has been integrally linked to guard cell physiology for decades, but has received little attention as a regulator of mesophyll cell biology. It is also important to note that this conclusion could not have arisen from extant physiological data, since to date no leaf responses to ABA have been evaluated at the physiological level in the G-protein mutant lines.

Although both guard cells and leaves do exhibit ABA–G-protein co-regulation of gene expression, the two systems favor different co-regulatory modes. For example, it is striking that at the transcriptome level, the classical II mechanism (**not GPA1 and AGB1**) is significantly more common in guard cells than in leaves, whereas the classical Ia and Ib mechanisms (**GPA1 and AGB1**) are possible in both tissue types, but more strongly so in leaves (Table V; see Supplementary information 14 for statistical analysis). In addition, the specific genes that are co-regulated by ABA and the G-protein differ between guard cells and leaves. Thus, among the 100 genes in guard cells and 549 genes in leaves that are ABA–G-protein co-regulated in some manner, only five genes (At1g30290, At4g29750, At1g31150, At1g74720, and At1g11210) appear in both guard cell and leaf data sets, in contrast to 106 ABA-only regulated genes common to both systems.

## Does regulation at the level of the transcriptome recapitulate previously observed hypersensitivity or hyposensitivity of developmental and dynamic transient responses to ABA in G-protein mutants?

Transcriptome level responses to ABA show some interesting differences compared with the existing physiological literature on ABA sensitivity of G-protein mutants (Wang *et al*, 2001; Pandey *et al*, 2006). For example, whereas ABA inhibition of stomatal opening and ion channel regulation in guard cells are ABA hyposensitive in *gpa1*, *agb1*, and *agb1 gpa1* double mutants (Wang *et al*, 2001; Fan *et al*, 2008), our analysis shows that ABA hyposensitivity of the mutant guard cell transcriptome, at least at the time point sampled here, is not as strong as the hyposensitivity evidenced by these rapid responses to ABA, possibly indicating differences between protein level and transcriptionally mediated responses. Further, a set of guard cells transcripts that shows ABA hypersensitivity in the G-protein mutants is observed. There is precedence for the same guard cell signaling element functioning in both ABA hyposensitivity and ABA hypersensitivity; for example, over-expression in guard cells of an inositol polyphosphate 5-phosphatase, which terminates inositol phosphate signaling, results in hyposensitivity to ABA inhibition of stomatal opening and hypersensitivity to ABA promotion of stomatal closure (Perera *et al*, 2008). Nevertheless, transcriptome results do not necessarily mean that ABA hypersensitivity of a portion of the transcriptome results in physiological outcomes that are ABA hypersensitive as, for example, the ABA-hypersensitive genes could encode elements that function in repression of ABA response. In either case, these results do indicate that a diversity of signaling pathways exists between the G-protein and gene regulation in guard cells. Interestingly, the majority of genes that show ABA hypersensitivity in the *gpa1* mutant (50/56) are governed by the classical II mechanism and the majority of genes that show equal hyposensitivity in all mutants (19/22) are governed by the classical I mechanisms.

Another novel observation is the strong ABA hyposensitivity of the G-protein-regulated mature leaf transcriptome, with all mutants showing equal ABA hyposensitivity. Previously, seed germination, seedling growth and gene expression were shown to be hypersensitive to ABA in these G-protein mutants (Pandey *et al*, 2006), leading to the hypothesis that ABA

hyposensitivity might be a distinctive characteristic found exclusively in the highly specialized guard cells, where ABA signaling has an integral function in cellular function. This hypothesis is not supported by our microarray data, which instead suggest the alternative hypothesis that G-protein involvement in ABA signaling is defined at the cell or organ level.

## Functional categories in G-protein regulatory modes

If the biological interpretations indicated by our model are valid, then we would expect to find at least some level of correlation between these interpretations of our transcriptome data and results from previous physiological analyses of G-protein mutants. Indeed, several of the functional categories identified by our analysis have been implicated in previous physiological analyses of G-protein mutants, providing validation to the biological interpretation of our results. Three functional categories related to plant pathogen response appear in guard cell regulatory modes, consistent with known roles of the G-protein in plant pathogen response (Trusov *et al*, 2006). For example, the functional category 'disease, virulence, and defense' is enriched in the regulatory mode $B_{11}$(ABA, $A_5$)=not (not GPA1 and AGB1)=GPA1 or not AGB1 in guard cells. Specifically, *At1G63750, At1G72520, LCR69, NHL10, NHL1,* and *AtPCB* are all genes implicated in defense against pathogens. This regulatory mode is consistent with a regulatory mechanism in which the Gβγ subunit, freed from its association with Gα, downregulates transcript levels of these defense genes, and is also consistent with our earlier physiological observations that *gpa1* guard cells are hyposensitive to the bacterial elicitor molecule, flg22 (Zhang *et al*, 2008b). Given this information, we might speculate that, in guard cells, ligands coupling to GPCRs upstream of the heterotrimer may be, or include, pathogen-derived or -dependent molecules, that is elicitors. Associated with the rule $A_6$=AGB1 in guard cells are the significant functional categories 'de/phosphorylation' and 'phosphate metabolism'. The latter category contains nine kinases and two genes encoding PP2C-type protein phosphatases. A number of kinases and PP2C phosphatases already have been shown to be integral to guard cell signaling pathways (Gosti *et al*, 1999; Li *et al*, 2000; Merlot *et al*, 2001; Mustilli *et al*, 2002; Yoshida *et al*, 2002; Ma *et al*, 2009; Park *et al*, 2009; Rubio *et al*, 2009); our results implicate new candidate proteins of this type. In guard cells, we also find that genes associated with metabolism of aromatic compounds are components of the ABA–G-protein co-regulated transcriptome. Although not yet studied in guard cells, there are several reports at the whole seedling level that implicate G-proteins in the regulation of aromatic amino-acid synthesis (Warpeha *et al*, 2006, 2008).

In leaves, we find that the functional category 'chloroplast' is highly significant, consistent with earlier observations linking GPA1 to chloroplast development (Zhang *et al*, 2009). Interestingly, the chloroplast protein THF1 (Zhang *et al*, 2009), which has been shown to physically interact with GPA1 (Huang *et al*, 2006), appears as one of the regulated transcripts in this category. Representation of functional categories related to signaling ('calcium binding,' 'cellular

signaling') is also consistent with the known signaling roles of G-proteins. The category 'nucleotide/nucleoside binding' is typified by kinases, ATPases, and ATP-binding G-proteins, linking G-proteins to well-established secondary messengers and transport functions. Thus, our transcriptome analyses are consistent with previously identified biological roles of G-proteins in Arabidopsis while also providing specific new gene targets for investigation.

One point to note is that, considering the data as a whole, the numbers of genes regulated by the **ABA-only** mode (713 in guard cells and 403 in leaves) is much greater than the numbers of genes regulated by G-protein-only modes (28 genes in guard cells and only At3G12730 in leaves, excluding GPA1 and AGB1 themselves). These results, along with the observation that transcription factors are not found to be overrepresented among **G-protein-only** or **G-protein–ABA co-regulated** genes (data not shown), suggests that the phenotypes of G-protein mutants do not arise due to a 'resetting' of the entire cellular profile but rather are likely to stem from disruption of specific G-protein-coupled signaling events.

In this study, we generate transcriptome data from stomatal guard cells and rosette leaves and investigate G-protein signaling in these two systems. There may be a cellular complexity difference between guard cell and leaf samples, as guard cells are a single cell type, whereas leaves are composed of several types of cells. However, the majority of the cell types contributing to the leaf microarray are mesophyll cells; guard cells and other non-mesophyll cell types comprise a very small percentage of the leaf. Our transcriptome measurements confirm that minor cell types such as guard cells do not significantly influence the leaf microarray. For example, in our wild-type transcriptome data sets, 361 transcripts were detectable in all three guard cell chip hybridizations but were not present in any of the three leaf chip hybridizations. In particular, some genes that are very highly expressed in Col guard cells, for example *ATHSP23.6-MITO* (AT4G25200), *ATATH4* (AT3G47760), *DDF1* (AT1G12610), *EPF1* (AT2G20875), *SP1L2* (AT1G69230), and *AtSIP1* (AT1G55740) are not present in any of the Col leaf microarrays.

In addition, in this study, we obtain leaves and guard cells at one developmental time point and measure gene expression at one time point after ABA treatment. Although we acknowledge these caveats, the former is not an issue as germination and leaf development throughout the vegetative stage is quite synchronous among all four genotypes under our conditions (Nilson and Assmann, 2010). For the latter, by our 3-h time point, the first (rapid) phase of guard cell response will have reached steady state, based on earlier studies of stomatal aperture response to ABA (Israelsson *et al*, 2006). Our study identifies key genes for which kinetics of gene expression can be assessed in more detail in subsequent studies.

## Relevance of our Boolean framework

Although Boolean networks have been used previously to extract gene relationships from time course expression profiles (Akutsu *et al*, 1999; Martin *et al*, 2007), this is, to our knowledge, the first study using systematic Boolean rules to deduce regulatory mechanisms based on analysis of mutant transcriptomes. The microarray studies in this work offer an

unprecedented complexity by studying four genotypes and two signaling conditions, in two different tissue types. With such complex data sets, pairwise differential expression analysis usually leads to results that are difficult to interpret because of the combinatorial complexity of regulation. The novelty of our method is that we provide an intuitive and reasonable way to analyze mutant expression data, which not only determines gene groups/modules, but also indicates the regulatory mechanisms connecting the genes in each group to the G-protein and/or to ABA. This makes our method very helpful for analyzing signal transduction pathways given individual signaling components with known properties, for example a transcription factor known to vary in phosphorylation status.

When associating genes with regulatory modes, we avoid biases and binarization of the expression values by correlating differential expression patterns instead of correlating absolute expression profiles. We use a systematic pairwise comparison procedure (Figure 2) to construct the differential expression vectors of genes. Not all of these pairwise comparisons are independent or orthogonal, but these comparisons determine a vector uniquely corresponding to a Boolean function from a comprehensive perspective and give a reliable result. We also tested the use of orthogonal contrasts to construct the differential expression patterns of genes, and obtained similar results (Supplementary information 15).

Clustering methods comprise another approach commonly used to predict co-regulated genes in microarray data sets (Thalamuthu *et al*, 2006; Liu *et al*, 2008). These methods are most useful when only rough knowledge such as upregulation or downregulation is expected as a result. The clusters do not offer information on the biological regulatory modes, whereas our method provides a putative mechanism for each group of co-regulated genes. In addition, clustering methods cannot distinguish the extent of a regulatory effect on a gene. In our method, each gene has correlation scores indicating the strength of the gene's association with each regulatory mode. We have tested the popular *k*-means clustering method to analyze our mutant expression data (Supplementary information 16) and find that most of the clusters do not have coherent expression patterns like ours (Supplementary Table S8). Moreover, these clusters lack biological interpretations such as those provided by our models.

In conclusion, our application of a Boolean framework for the analysis of microarray data sets has discovered new mechanisms of G-protein and hormonal control. We posit that application of this approach to transcriptomic data sets from other systems will similarly provide new perspectives regarding other switch-like signal transduction mechanisms, such as regulation by reversible post-translational events or by combinatorial association of transcription factors.

# Materials and methods

## Plant materials, ABA treatment, and microarray experiments

All mutants are in the Arabidopsis Col background and have been described earlier (Jones *et al*, 2003; Ullah *et al*, 2003; Chen *et al*, 2004). These are T-DNA insertional mutants that fail to produce full-length *GPA1* or *AGB1* transcripts; *gpa1-4* is also confirmed as a null mutation

at the protein level (Fan *et al*, 2008). Microarray data were generated from four genotypes (wild type, *gpa1-4* mutant, *agb1-2* mutant, and *agb1-2 gpa1-4* double mutant) with or without ABA treatment. Arabidopsis plants were grown in growth chambers with an 8 h light/16 h dark, 20°C/18°C cycle, with light intensity of $120 \, \mu mol \, m^{-2} \, s^{-1}$. Three hundred Arabidopsis leaves excised from 60 to 70 five-week-old plants were used as the starting material for each guard cell microarray (see Supplementary information 2 for plant growth details and guard cell isolation protocol). Ten mature leaves taken from three to four plants grown side-by-side with the plants for guard cell isolation were used for each leaf sample. Excised leaf and isolated guard cell samples were treated with ABA (50 μM) or EtOH (solvent control) for 3 h. RNA was isolated from each sample using Trizol reagent (Invitrogen). Trizol-isolated RNA was further purified using the Plant RNeasy kit (Qiagen, Valencia, CA). RNA samples exhibiting a 25S/18S ratio >1.4 and no significant degradation in Bioanalyzer profiles (Agilent Technology, Palo Alto, CA) were used for cDNA and then cRNA synthesis, followed by hybridization to Affymetrix ATH1 'whole genome' chips, all according to standard Affymetrix protocols. For each type of sample (guard cells or leaves), three independent biological replicates were performed, resulting in a total of 48 microarray hybridizations (2 sample types × 4 genotypes × two treatments × 3 replicates).

## Microarray data processing and validation

The transcriptome data reported in this paper have been deposited in the Gene Expression Omnibus database (http://www.ncbi.nlm.nih.gov/geo) with accession no. GSE19520. We applied several different approaches to assess quality control and the overall quality of the hybridizations is very high (Supplementary information 17). The average pairwise correlation among all biological replicates is 0.982, indicating that the variation arising from biological replicates is very small. The widely used RMA method implemented in the affy package in the bioconductor project was used to normalize the data for all probe sets using default parameters (Gautier *et al*, 2004; Gentleman *et al*, 2004). The RMA does not provide a way to calculate present calls and absent calls, so GeneChip suite 5.0 algorithm implemented in the affy package was used to obtain presence/absence calls for each probe set in each independent sample with default parameters. In total, there are 22 810 probe sets on the Affymetrix ATH1 Arabidopsis chips. We eliminated only those genes that were absent in all 24 samples of a given tissue type (a very loose filtering); this left 17 581 probes from guard cell samples and 17 827 probes from leaf samples for further analysis. The subsequent analysis of variance (ANOVA) step discards those genes that show no differential expression in any of our microarray comparisons.

Data obtained from the microarray hybridizations were confirmed by real-time quantitative reverse transcriptase PCR (QPCR) of selected genes (Supplementary information 18). For each gene, the differential expression pattern of all genotype × treatment combinations was compared between microarray data and QPCR data (8 comparisons per gene, for a total of 64 comparisons). Microarray and QPCR patterns of gene differential expression exhibit an 86% (=55/64) match (see Supplementary Figure S16).

## Boolean model of G-protein regulatory modes

We represent GPA1 and AGB1 by Boolean variables that can have two states: 1 denoting 'on' (not knocked out) and 0 denoting 'off' (knocked out). Each of the four genotypes considered in our study corresponds to one of the $2^2=4$ possible combinations of the states of GPA1 and AGB1, from GPA1=0 and AGB1=0 for the *agb1 gpa1* double mutant to GPA1=1 and AGB1=1 for wild type. The regulatory modes of the G-protein are modeled by a Boolean function A(GPA1, AGB1). There is a total of $2^4=16$ possible Boolean functions fitting A(GPA1, AGB1), which we denote as $A_i$(GPA1, AGB1), $i=1,2,\ldots,16$. Except $A_1$ and $A_{16}$, each of these corresponds to one regulatory mode of the G-protein and reflects the activity of a G-protein-regulated mediator that further regulates the expression level of a target gene. All possible regulatory modes of the G-protein and the associated expression patterns of their potential target genes can be found in Supplementary Table S1.

## Boolean model of G-protein–ABA co-regulatory modes

We also represent ABA by a Boolean variable with 1 indicating the presence of ABA (i.e. ABA treatment) and 0 representing the absence of ABA (i.e. solvent control). We assume that co-regulation by the G-protein and ABA can be described as F(ABA, GPA1, AGB1) = $C_{ABA}$ + B(ABA, GPA1, AGB1), where $C_{ABA}$ is a constant representing a G-protein-independent regulatory effect of ABA on a given gene, and B(ABA, GPA1, AGB1) is a Boolean function representing the possible combinatorial regulation of ABA and the G-protein. A value for $C_{ABA}$ that is not vanishingly small indicates that there is a G-protein-independent signaling pathway by which ABA regulates the gene. Although a general B(ABA, GPA1, AGB1) function allows a variety of combinatorial regulatory schemes by individual G-protein subunits together with ABA, we have not found evidence of ABA participating in contrasting combinatorial regulation with individual G-protein subunits (see Supplementary information 5 for a detailed analysis). Thus, in the following analysis, we use the simplified form B(ABA, A(GPA1, AGB1)) that excludes these cases.

In total, there are $2^4 = 16$ Boolean functions fitting the B(ABA, A(GPA1, AGB1)) form, denoted as $B_1$, $B_2$,..., $B_{16}$. Each of these represents a set of Boolean rules due to the fact that A(GPA1, AGB1) is already a Boolean function. Because of logical equivalence, in total, there are 72 non-trivial unique Boolean functions that we can use to distinguish between (co)regulatory modes of the G-protein and ABA (see Supplementary information 6 for the derivation of these Boolean functions). Together with $C_{ABA}$, F(ABA, GPA1, AGB1) is able to identify 142 (i.e. $(72-2) \times 2 + 2$) regulatory modes, which can categorize the regulation by the G-protein and ABA into five types: **ABA-only** regulation ($B_4$, $B_{13}$), **G-protein-only** regulation ($B_6$ (ABA, $A_i$) = $B_{11}$ (ABA, $A_{17-i}$), $i=2,...,15$, combined with a non-significant $C_{ABA}$), G-protein–ABA **additive** regulation ($B_6$ (ABA, $A_i$) = $B_{11}$ (ABA, $A_{17-i}$), $i=2,...,15$, combined with a significant $C_{ABA}$), ABA–G-protein **combinatorial** regulation ($B_2$(ABA, $A_i$) = $B_3$(ABA, $A_{17-i}$), $B_5$(ABA, $A_i$) = $B_9$(ABA, $A_{17-i}$), $B_8$(ABA, $A_i$) = $B_{12}$(ABA, $A_{17-i}$), $B_{14}$(ABA, $A_i$) = $B_{15}$(ABA, $A_{17-i}$), $i=2,...,15$, combined with a non-significant $C_{ABA}$) and ABA–G-protein **mixed** regulation ($B_2$(ABA, $A_i$) = $B_3$(ABA, $A_{17-i}$), $B_5$(ABA, $A_i$) = $B_9$(ABA, $A_{17-i}$), $B_8$(ABA, $A_i$) = $B_{12}$(ABA, $A_{17-i}$), $B_{14}$(ABA, $A_i$) = $B_{15}$(ABA, $A_{17-i}$), $i=2,...,15$, combined with a significant $C_{ABA}$), which further allow us to examine the representation of the nine proposed G-protein and ABA signaling pathways (Supplementary Figure S1).

## Association of genes with regulatory modes

As a first step, we use ANOVA (Sahai and Agell, 2000) to identify the genes that are differentially expressed with respect to genotypes or ABA. Within the control or ABA treatment categories, the expression of each gene is affected merely by genotypes, so we perform one-way ANOVA and denote the corresponding $P$-value as $P_{geno}$. The expression of each gene in all samples is affected by both genotypes and ABA. To extract the ABA effect, we use two-way ANOVA and denote the ABA-associated $P$-value as $P_{ABA}$.

For association of genes with G-protein regulatory modes, we use $P_{geno} < 0.01$ for each individual treatment × tissue combination. For association of genes with G-protein–ABA co-regulatory modes, we designate $P > 0.05$ as no differential expression and $P < 0.02$ as differential expression. Specifically, there are 6975 genes in guard cells and 6299 genes in leaves that have $P_{ABA} > 0.05$ and $P_{geno} > 0.05$ (no differential expression with respect to genotypes and also no differential expression with respect to ABA) and these genes are not considered further. Those genes that have no differential expression with respect to genotypes but are differentially expressed with respect to ABA ($P_{ABA} < 0.02$ and $P_{geno} > 0.05$ for both ABA=0 and ABA=1) are candidates for ABA-only regulation (Figure 1A). For those genes with differential expression in at least one condition of ABA=0 and ABA=1, we determine and subtract $C_{ABA}$ from the expression values of the genes in the condition (presence or absence of ABA) corresponding to the higher level. We choose $C_{ABA} = 0.6$ as the lower threshold of significance (Supplementary information 7). Those genes that have genotype-dependent differential expression in both ABA=0 and

ABA=1 ($P_{geno} < 0.02$ for both ABA=0 and ABA=1) are candidates for G-protein-only regulation (if their $C_{ABA}$ is not significant) (Figure 1B) or G-protein–ABA additive regulation (if their $C_{ABA}$ is significant) (Figure 1C). Those genes that exhibit differential expression with respect to genotypes for ABA=0 or ABA=1, but not both ($P_{geno} < 0.02$ in one case and $P_{geno} > 0.05$ in the other case) are candidates for ABA–G-protein combinatorial regulation (if their $C_{ABA}$ is not significant) (Figure 1D) or ABA–G-protein mixed regulation (if their $C_{ABA}$ is significant) (Supplementary Figure S1E).

Each regulatory mode of the G-protein and ABA corresponds to a unique three-valued vector, which we call an idealized differential expression pattern, through comparison of the gene expression values between pairs of genotypes or within the same genotype in the absence or presence of ABA (Figure 2; Supplementary information 8). The real differential expression pattern of a gene is constructed using a signal-to-noise ratio metric (Golub *et al*, 1999). We introduce a measure to quantify the correlation between the real differential expression pattern of a gene with a Boolean function's three-valued vector (see Supplementary information 8). Each gene is associated with that Boolean function for which it has the maximum correlation score. We classify a gene as regulated by a particular mode if its correlation score is higher than a threshold.

## False discovery rates and statistical significance evaluation

Permutation tests are used to determine the statistical significance of each gene's association with a given function, and to estimate the false discovery rate (FDR) in putative gene groups of particular type (see Supplementary information 8). The correlation threshold is chosen in such a way to control the FDR at an acceptable level (see Supplementary information 9 and 10). Specifically, for association of genes with G-protein regulatory modes, in all four data sets, differentially expressed genes are associated with Boolean functions A(GPA1, AGB1) if the correlation between their differential expression patterns and the idealized pattern exceeds the threshold $R_0 = 1.1$, which controls the FDR within 0.02 (see Supplementary Figure S6). For association of genes with G-protein–ABA co-regulatory modes the correlation threshold $R_0 = 0.9$ is used, which controls the FDR of ABA–G-protein combinatorially or mixed regulated genes within 0.05 and the FDR of G-protein-only or G-protein–ABA additively regulated genes within 0.005 (Supplementary Figures S7 and S8). As the number of ABA-only regulated genes is large, we set 2.0 as the threshold of ABA-only regulated genes, which controls the FDR within 0.0005 (Supplementary Figure S9). These thresholds and FDRs apply to both guard cell data and leaf data.

## Supplementary information

Supplementary information is available at the *Molecular Systems Biology* website (http://www.nature.com/msb).

## Conflict of interest

The authors declare that they have no conflict of interest.

## References

Acharya BR, Assmann SM (2009) Hormone interactions in stomatal function. *Plant Mol Biol* **69:** 451–462

Akutsu T, Miyano S, Kuhara S (1999) Identification of genetic networks from a small number of gene expression patterns under the Boolean network model. *Pac Symp Biocomput* **4:** 17–28

Assmann SM (2005) G proteins Go green: a plant G protein signaling FAQ sheet. *Science* **310:** 71–73

Chen JG, Gao Y, Jones AM (2006) Differential roles of Arabidopsis heterotrimeric G-protein subunits in modulating cell division in roots. *Plant Physiol* **141:** 887–897

Chen JG, Pandey S, Huang J, Alonso JM, Ecker JR, Assmann SM, Jones AM (2004) GCR1 can act independently of heterotrimeric G-protein in response to brassinosteroids and gibberellins in Arabidopsis seed germination. *Plant Physiol* **135:** 907–915

Coursol S, Fan LM, Le Stunff H, Spiegel S, Gilroy S, Assmann SM (2003) Sphingolipid signalling in Arabidopsis guard cells involves heterotrimeric G proteins. *Nature* **423:** 651–654

Digby GJ, Sethi PR, Lambert NA (2008) Differential dissociation of G protein heterotrimers. *J Physiol* **586:** 3325–3335

Dupre DJ, Robitaille M, Rebois RV, Hebert TE (2009) The role of Gβγ subunits in the organization, assembly, and function of GPCR signaling complexes. *Annu Rev Pharmacol Toxicol* **49:** 31–56

Fan LM, Zhang W, Chen JG, Taylor JP, Jones AM, Assmann SM (2008) Abscisic acid regulation of guard-cell K$^+$ and anion channels in Gβ- and RGS-deficient Arabidopsis lines. *Proc Natl Acad Sci USA* **105:** 8476–8481

Frank M, Thumer L, Lohse MJ, Bunemann M (2005) G Protein activation without subunit dissociation depends on a Gα(i)-specific region. *J Biol Chem* **280:** 24584–24590

Finkelstein RR, Gampala SS, Rock CD (2002) Abscisic acid signaling in seeds and seedlings. *Plant Cell* **14** (Suppl)**:** S15–S45

Gautier L, Cope L, Bolstad BM, Irizarry RA (2004) affy—Analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics* **20:** 307–315

Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, Ellis B, Gautier L, Ge Y, Gentry J, Hornik K, Hothorn T, Huber W, Iacus S, Irizarry R, Leisch F, Li C, Maechler M, Rossini AJ, Sawitzki G *et al* (2004) Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol* **5:** R80

Golub TR, Slonim DK, Tamayo P, Huard C, Gaasenbeek M, Mesirov JP, Coller H, Loh ML, Downing JR, Caligiuri MA, Bloomfield CD, Lander ES (1999) Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science* **286:** 531–537

Gookin TE, Kim J, Assmann SM (2008) Whole proteome identification of plant candidate G-protein coupled receptors in Arabidopsis, rice, and poplar: computational prediction and *in-vivo* protein coupling. *Genome Biol* **9:** R120

Gosti F, Beaudoin N, Serizet C, Webb AA, Vartanian N, Giraudat J (1999) ABI1 protein phosphatase 2C is a negative regulator of abscisic acid signaling. *Plant Cell* **11:** 1897–1910

Huang J, Taylor JP, Chen JG, Uhrig JF, Schnell DJ, Nakagawa T, Korth KL, Jones AM (2006) The plastid protein THYLAKOID FORMATION1 and the plasma membrane G-protein GPA1 interact in a novel sugar-signaling mechanism in Arabidopsis. *Plant Cell* **18:** 1226–1238

Israelsson M, Siegel RS, Young J, Hashimoto M, Iba K, Schroeder JI (2006) Guard cell ABA and CO$_2$ signaling network updates and Ca$^{2+}$ sensor priming hypothesis. *Curr Opin Plant Biol* **9:** 654–663

Johnston CA, Taylor JP, Gao Y, Kimple AJ, Grigston JC, Chen JG, Siderovski DP, Jones AM, Willard FS (2007) GTPase acceleration as the rate-limiting step in Arabidopsis G protein-coupled sugar signaling. *Proc Natl Acad Sci USA* **104:** 17317–17322

Jones AM, Assmann SM (2004) Plants: the latest model system for G-protein research. *EMBO Rep* **5:** 572–578

Jones AM, Ecker JR, Chen JG (2003) A reevaluation of the role of the heterotrimeric G protein in coupling light responses in Arabidopsis. *Plant Physiol* **131:** 1623–1627

Klein S, Reuveni H, Levitzki A (2000) Signal transduction by a nondissociable heterotrimeric yeast G protein. *Proc Natl Acad Sci USA* **97:** 3219–3223

Leaney JL, Tinker A (2000) The role of members of the pertussis toxin-sensitive family of G proteins in coupling receptors to the activation of the G protein-gated inwardly rectifying potassium channel. *Proc Natl Acad Sci USA* **97:** 5651–5656

Lease KA, Wen J, Li J, Doke JT, Liscum E, Walker JC (2001) A mutant Arabidopsis heterotrimeric G-protein β subunit affects leaf, flower, and fruit development. *Plant Cell* **13:** 2631–2641

Leonhardt N, Kwak JM, Robert N, Waner D, Leonhardt G, Schroeder JI (2004) Microarray expression analyses of Arabidopsis guard cells and isolation of a recessive abscisic acid hypersensitive protein phosphatase 2C mutant. *Plant Cell* **16:** 596–615

Leung J, Giraudat J (1998) Abscisic acid signal transduction. *Annu Rev Plant Physiol Plant Mol Biol* **49:** 199–222

Li J, Wang XQ, Watson MB, Assmann SM (2000) Regulation of abscisic acid-induced stomatal closure and anion channels by guard cell AAPK kinase. *Science* **287:** 300–303

Liu X, Lee S-C, Casella G, Peter GF (2008) Assessing agreement of clustering methods with gene expression microarray data. *Comput Stat Data Anal* **52:** 5356–5366

Ma Y, Szostkiewicz I, Korte A, Moes D, Yang Y, Christmann A, Grill E (2009) Regulators of PP2C phosphatase activity function as abscisic acid sensors. *Science* **324:** 1064–1068

Marrari Y, Crouthamel M, Irannejad R, Wedegaertner PB (2007) Assembly and trafficking of heterotrimeric G proteins. *Biochemistry* **46:** 7665–7677

Martin S, Zhang Z, Martino A, Faulon JL (2007) Boolean dynamics of genetic regulatory networks inferred from microarray time series data. *Bioinformatics* **23:** 866–874

McCudden CR, Hains MD, Kimple RJ, Siderovski DP, Willard FS (2005) G-protein signaling: back to the future. *Cell Mol Life Sci* **62:** 551–577

Melotto M, Underwood W, He SY (2008) Role of stomata in plant innate immunity and foliar bacterial diseases. *Annu Rev Phytopathol* **46:** 101–122

Merlot S, Gosti F, Guerrier D, Vavasseur A, Giraudat J (2001) The ABI1 and ABI2 protein phosphatases 2C act in a negative feedback regulatory loop of the abscisic acid signalling pathway. *Plant J* **25:** 295–303

Mewes HW, Frishman D, Mayer KF, Munsterkotter M, Noubibou O, Pagel P, Rattei T, Oesterheld M, Ruepp A, Stumpflen V (2006) MIPS: analysis and annotation of proteins from whole genomes in 2005. *Nucleic Acids Res* **34:** D169–D172

Moriyama EN, Strope PK, Opiyo SO, Chen Z, Jones AM (2006) Mining the *Arabidopsis thaliana* genome for highly-divergent seven transmembrane receptors. *Genome Biol* **7:** R96

Morris AJ, Malbon CC (1999) Physiological regulation of G protein-linked signaling. *Physiol Rev* **79:** 1373–1430

Mustilli AC, Merlot S, Vavasseur A, Fenzi F, Giraudat J (2002) Arabidopsis OST1 protein kinase mediates the regulation of stomatal aperture by abscisic acid and acts upstream of reactive oxygen species production. *Plant Cell* **14:** 3089–3099

Nilson SE, Assmann SM (2010) Heterotrimeric G proteins regulate reproductive trait plasticity in response to water availability. *New Phytol* **185:** 734–746

Okamoto H, Gobel C, Capper RG, Saunders N, Feussner I, Knight MR (2009) The α-subunit of the heterotrimeric G-protein affects jasmonate responses in *Arabidopsis thaliana*. *J Exp Bot* **60:** 1991–2003

Oldham WM, Hamm HE (2008) Heterotrimeric G protein activation by G-protein-coupled receptors. *Nat Rev Mol Cell Biol* **9:** 60–71

Pandey S, Chen JG, Jones AM, Assmann SM (2006) G-protein complex mutants are hypersensitive to abscisic acid regulation of germination and postgermination development. *Plant Physiol* **141:** 243–256

Pandey S, Monshausen GB, Ding L, Assmann SM (2008) Regulation of root-wave response by extra large and conventional G proteins in *Arabidopsis thaliana*. *Plant J* **55:** 311–322

Pandey S, Nelson DC, Assmann SM (2009) Two novel GPCR-type G proteins are abscisic acid receptors in Arabidopsis. *Cell* **136:** 136–148

Park SY, Fung P, Nishimura N, Jensen DR, Fujii H, Zhao Y, Lumba S, Santiago J, Rodrigues A, Chow TF, Alfred SE, Bonetta D, Finkelstein R, Provart NJ, Desveaux D, Rodriguez PL, McCourt P, Zhu JK, Schroeder JI, Volkman BF *et al* (2009) Abscisic acid inhibits type 2C protein phosphatases via the PYR/PYL family of START proteins. *Science* **324:** 1068–1071

Pavlidis P, Noble WS (2003) Matrix2png: a utility for visualizing matrix data. *Bioinformatics* **19:** 295–296

Perera IY, Hung CY, Moore CD, Stevenson-Paulik J, Boss WF (2008) Transgenic Arabidopsis plants expressing the type 1 inositol 5-phosphatase exhibit increased drought tolerance and altered abscisic acid signaling. *Plant Cell* **20:** 2876–2893

Perfus-Barbeoch L, Jones AM, Assmann SM (2004) Plant heterotrimeric G protein function: insights from Arabidopsis and rice mutants. *Curr Opin Plant Biol* **7:** 719–731

Roelfsema MR, Hedrich R (2005) In the light of stomatal opening: new insights into 'the Watergate'. *New Phytol* **167:** 665–691

Rubio S, Rodrigues A, Saez A, Dizon MB, Galle A, Kim TH, Santiago J, Flexas J, Schroeder JI, Rodriguez PL (2009) Triple loss of function of protein phosphatases type 2C leads to partial constitutive response to endogenous abscisic acid. *Plant Physiol* **150:** 1345–1355

Sahai H, Agell MI (2000) *The Analysis of Variance*. Boston: Birkhauser

Shimazaki K, Doi M, Assmann SM, Kinoshita T (2007) Light regulation of stomatal movement. *Annu Rev Plant Biol* **58:** 219–247

Temple BR, Jones AM (2007) The plant heterotrimeric G-protein complex. *Annu Rev Plant Biol* **58:** 249–266

Thalamuthu A, Mukhopadhyay I, Zheng X, Tseng GC (2006) Evaluation and comparison of gene clustering methods in microarray analysis. *Bioinformatics* **22:** 2405–2412

Trusov Y, Rookes JE, Chakravorty D, Armour D, Schenk PM, Botella JR (2006) Heterotrimeric G proteins facilitate Arabidopsis resistance to necrotrophic pathogens and are involved in jasmonate signaling. *Plant Physiol* **140:** 210–220

Ullah H, Chen JG, Temple B, Boyes DC, Alonso JM, Davis KR, Ecker JR, Jones AM (2003) The β-subunit of the Arabidopsis G protein negatively regulates auxin-induced cell division and affects multiple developmental processes. *Plant Cell* **15:** 393–409

Wang S, Narendra S, Fedoroff N (2007) Heterotrimeric G protein signaling in the Arabidopsis unfolded protein response. *Proc Natl Acad Sci USA* **104:** 3817–3822

Wang XQ, Ullah H, Jones AM, Assmann SM (2001) G protein regulation of ion channels and abscisic acid signaling in Arabidopsis guard cells. *Science* **292:** 2070–2072

Warpeha KM, Gibbons J, Carol A, Slusser J, Tree R, Durham W, Kaufman LS (2008) Adequate phenylalanine synthesis mediated by G protein is critical for protection from UV radiation damage in young etiolated *Arabidopsis thaliana* seedlings. *Plant Cell Environ* **31:** 1756–1770

Warpeha KM, Lateef SS, Lapik Y, Anderson M, Lee BS, Kaufman LS (2006) G-protein-coupled receptor 1, G-protein Gα-subunit 1, and prephenate dehydratase 1 are required for blue light-induced production of phenylalanine in etiolated Arabidopsis. *Plant Physiol* **140:** 844–855

Yoshida R, Hobo T, Ichimura K, Mizoguchi T, Takahashi F, Aronso J, Ecker JR, Shinozaki K (2002) ABA-activated SnRK2 protein kinase is required for dehydration stress signaling in Arabidopsis. *Plant Cell Physiol* **43:** 1473–1483

Zhang L, Hu G, Cheng Y, Huang J (2008a) Heterotrimeric G protein α and β subunits antagonistically modulate stomatal density in *Arabidopsis thaliana*. *Dev Biol* **324:** 68–75

Zhang L, Wei Q, Wu W, Cheng Y, Hu G, Hu F, Sun Y, Zhu Y, Sakamoto W, Huang J (2009) Activation of the heterotrimeric G protein α-subunit GPA1 suppresses the ftsh-mediated inhibition of chloroplast development in Arabidopsis. *Plant J* **58:** 1041–1053

Zhang W, He SY, Assmann SM (2008b) The plant innate immunity response in stomatal guard cells invokes G-protein-dependent ion channel regulation. *Plant J* **56:** 984–996