Arabidopsis G-protein interactome reveals connections to cell wall carbohydrates and morphogenesis

Karsten Klopffleisch, Nguyen Phan, Kelsey Augustin, Robert S Bayne, Katherine S Booker, Jose R Botella, Nicholas C Carpita, Tyrell Carr, Jin-Gui Chen, Thomas Ryan Cooke, Arwen Frick-Cheng, Erin J Friedman, Brandon Fulk, Michael G Hahn, Kun Jiang, Lucia Jorda, Lydia Kruppe, Chenggang Liu, Justine Lorek, Maureen C McCann, Antonio Molina, Etsuko N Moriyama, M Shahid Mukhtar, Yashwanti Mudgil, Sivakumar Pattathil, John Schwarz, Mr. Steven Seta, Matthew Tan, Ulrike Temp, Yuri Trusov, Daisuke Urano, Bastian Welter, Jing Yang, Ralph Panstruga, Joachim F Uhrig and Alan M Jones

Corresponding author: Alan Jones, University of North Carolina

Review timeline:

<table>
<thead>
<tr>
<th>Event</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>Submission date</td>
<td>02 May 2011</td>
</tr>
<tr>
<td>Editorial Decision</td>
<td>09 June 2011</td>
</tr>
<tr>
<td>Revision received</td>
<td>07 July 2011</td>
</tr>
<tr>
<td>Editorial Decision</td>
<td>05 August 2011</td>
</tr>
<tr>
<td>Revision received</td>
<td>15 August 2011</td>
</tr>
<tr>
<td>Editorial Decision</td>
<td>15 August 2011</td>
</tr>
<tr>
<td>Revision received</td>
<td>17 August 2011</td>
</tr>
<tr>
<td>Accepted</td>
<td>17 August 2011</td>
</tr>
</tbody>
</table>

Transaction Report:

(Note: With the exception of the correction of typographical or spelling errors that could be a source of ambiguity, letters and reports are not edited. The original formatting of letters and referee reports may not be reflected in this compilation.)

1st Editorial Decision 09 June 2011

Thank you again for submitting your work to Molecular Systems Biology. We have now heard back from the three referees whom we asked to evaluate your manuscript. As you will see from the reports below, the referees raise substantial concerns, which, I am afraid to say, preclude its publication in its present form.

In general, the first two reviewers found the G-protein-centered interaction network produced in this work, and its connection to cell wall synthesis, to be potentially interesting. They raise, however, some concerns regarding the yeast 2-hybrid data and the BiFC validation experiments. In particular, Reviewer #1 indicated clearly that additional control experiments were needed to support the interaction data, including tests to exclude typical artifacts like autoactivation, and a series of negative controls for the BiFC data.

The third reviewer's opinion was less positive, and this reviewer had some important concerns regarding the stomatal and cell wall experiments. While the editor feels that extensive exploration of the mechanisms linking G-protein signaling and the cell wall phenotypes would lie outside the scope of this short Report, it is essential that the link between G-protein signaling and these phenotypes is demonstrated convincing, especially since these experiments provide the main evidence supporting the functional relevance of the G-protein interaction network. As such, additional clarification and,
possibly, further experimentation, addressing this reviewer's technical concerns, appears to be required.

Overall, the reviewers had some more general concerns about the presentation (e.g. point #6 from Reviewer #1), and the editor notes that you may include up to one additional Figure in the main manuscript.

In addition, before submitting any revised work, we ask that you deposit the molecular interaction data with a member of the International Molecular Exchange Consortium (IMEx, http://www.imexconsortium.org), following the MIMIx recommendations (http://www.psidev.info/index.php?q=node/278).

If you feel you can satisfactorily deal with these points and those listed by the referees, you may wish to submit a revised version of your manuscript. Please attach a covering letter giving details of the way in which you have handled each of the points raised by the referees. A revised manuscript will be once again subject to review and you probably understand that we can give you no guarantee at this stage that the eventual outcome will be favorable.

*PLEASE NOTE* As part of the EMBO Publications transparent editorial process initiative (see http://www.nature.com/msb/journal/v6/n1/full/msb201072.html), Molecular Systems Biology now publishes online a Review Process File with each accepted manuscript. Please be aware that in the event of acceptance, your cover letter/point-by-point document will be included as part of this file, which will be available to the scientific community. More information about this initiative is available in our Instructions to Authors. If you have any questions about this initiative, please contact the editorial office (msb@embo.org).

Yours sincerely,

Editor - Molecular Systems Biology
msb@embo.org

---------------------------------------------------------------------------

REFEREE REPORTS
---------------------------------------------------------------------------

Reviewer #1 (Remarks to the Author):

The manuscript describes the mapping of a G-protein centered physical protein-protein interaction map for Arabidopsis thaliana. Starting from heterotromeric protein subunits, the authors perform extensive Y2H screening using cDNAs from multiple tissues. Proteins interacting with multiple baits were rescreeen to expand the network. For the final dataset, proteins interacting with at least two proteins were considered reliable and chose for further analysis.

For validation 63 interaction pairs involving 24 proteins were tested by BiFC in planta, of which 62 pairs scored positive. The cellular and subcellular localizations for these were analyzed. For coexpression analysis a positive gold standard and random proteins were used to calibrate the analysis. Then the co-expression correlation for all 544 pairs found to be interacting is calculated and found to be significant.

By manual inspection of the interactome the authors then hypothesize that G-proteins may be involved in Cell Wall synthesis. They proceed to show this involvement using a panel of biochemical analytical tools applied to at least 7 mutant lines.

The manuscript is very well written and addresses an important topic in plant systems analysis. Despite my overall favorable view of the manuscript, important questions need to be addressed.

1. For the Y2H analysis, the experimental section needs to expanded. Especially the well know artifact of autoactivation and spontaneous autoactivation need to be addressed. Additional experimental controls may need to be performed here.

2. Validation: I applaud the authors for doing an experimental validation that is unbiased by
previous knowledge (as computational analyses necessarily are) and to do so on a significant part of the network. However, the validation rate of >95% is extremely high and a reason for concern. It was previously shown using calibrated interaction assays (Braun et al., Nat Meth 2009), that different assays have distinct protein-interaction-detection profiles and under stringent conditions an orthogonal assay detects only <50% of interactions that were identified using a different assay. The concern here is that the BiFC assay is either done or scored with very promiscuous conditions and that the validation is therefore invalid.

To address this concern, I recommend taking a set of random protein pairs, that are not expected to interact (no less than 20), and test this under the very same conditions, to obtain an estimate of the false positive rate of BiFC.

3. I am also concerned about the fact that only pairs from the core-interactome are validated, whereas the manuscript describes a much larger network for a large part of which no estimation of the quality is provided. It is recommended to pick pairs for validation that either come to equal parts from the core and non-core dataset. Most appropriately interacting pairs should be picked at random from the 544 interactions and then both datasets will be represented to equal parts in the validation tests.

4. The statement that 'triangles are overrepresented in the network' is meaningless and misleading and should be removed as this is a consequence of the biased experimental design and filtering. It could be replaced by 'abundant'.

5. The numbers of co-expression correlated gene pairs in confusing. After some mulling of the sentence, I suspect that not all of the 544 pairs could be tested in all categories. However, this is not well explained and the ratio of correlated pairs versus tested pairs should be presented for each category. Further, why are 903 pairs tested when the interaction dataset only encompasses 544 protein pairs? This section needs to be completely reworked and more details on findings should be provided. It would be interesting, e.g. to see whether the pairs in a certain category have a higher co-expression correlation in appropriate datasets than pairs not in the category in the same transcription datasets.

6. For the cell wall findings an illustrative figure would be instructive. Potentially a cartoon of how different analyses fit together would help the reader. I am not entirely sure what is missing, but overall the figures accompanying this manuscript are less helpful than they could be. On the other hand, I am not entirely clear what the message of the network figures is and if it is necessary. Network representations can be very informative, however from the presented graphs I gather the message that the original bait proteins are most highly connected, which is hardly surprising and due to the experimental design. Perhaps the core network can be a panel of the last figure, and the cell wall synthesis proteins in the network could be highlighted to justify the authors investigation of this process.

Reviewer #2 (Remarks to the Author):

Overall, this manuscript is interesting. It presents a two hybrid analysis of proteins that interact with G proteins, initially in an unbiased manner, but at later points using a more biased approach. The authors do minor network analysis, and major in planta validation. They find an interesting novel interaction between G proteins and cell wall biosynthesis. Overall, the paper reads almost as two distinct studies - one with network analysis/coexpression - then jumps into a non-network analysis that does not use quantitatively derived network topology to characterize its biology. Perhaps this could be addressed by some rewriting and some additional statistical analysis of the biological functions of the nodes within the network.

Major Points:
1. Detected 544 interactions between 434 proteins - a comprehensive interactome

How many times was a particular interactor found in each library for the first screen? The authors state "from the total of 206 unique prey proteins from the first round of screenings, 13 secondary baits were chosen because they interacted with two or more primary baits". What makes these baits unique? The fact that they were found once? If so - then I would say that this network is not comprehensive. There should be a threshold for the number of times a prey is found (sufficient coverage) to interact in any given library before deciding on the secondary screen. What if there are effectors that interact with just a single bait protein. Does this make them any less functionally important? Is there support for this in the literature?
2. 45% of interactions were validated in planta

This is a particularly impressive rate of validation. What is the relevance of the expression patterns of these signals? Is there evidence in the literature that these expression patterns generally reflect the in vivo subcellular location of these genes? Transient transformation in tobacco seems a little artificial. More importantly, I would want to know if these gene pairs are at least expressed in the same context in at least one of the AtGenExpress arrays?

3. Co-expression analysis reveals that most are co-expressed (problems with null distribution)

I am concerned about the method of switching the pairs of the gold-standard proteins to create a null distribution. Do these conclusions hold when a set of genes are randomly chosen from the Arabidopsis genome? This particular null distribution isn't quite null since the gold standard positive set are also likely to be correlated in their expression. Their co-expression would be higher than that of the genome - they are likely being overly conservative, and aren't really testing if the co-expression levels are much higher than what would be expected by random chance. Perhaps even consider using the distribution of co-expression values from negatives obtained by the screening, and then compare the distribution of the positives from the screening?

Also - the authors mention co-expression within network motifs within this section as justification for this analysis. However - no analysis was performed to see if there is any significant structure within this network, and any such analysis would be unjustified given the experimental design for the choice of additional baits in the second screen. I would therefore remove this sentence, and I would look at the structure of the network more quantitatively.

4. Reveal a novel role for G proteins in regulating cell wall modification

The authors mention protein clustering on page 7. I see no evidence of protein clustering.

Is the presence of genes regulating the cell wall a significant observation? In a randomized network with the same structure, what would the distribution be of genes with a cell wall annotation?

The cell wall analysis is really striking, particularly because, to the best of my knowledge, this is the first report linking G protein signaling to the cell wall. The biochemical analysis is well performed, as is the clustering of the antibody epitopes on the cell wall. The difference between the root and the shoot is also interesting. Are both these genes expressed equally in the root or the shoot, or are they expressed in different cell types in the shoot and root that would potentially explain these differences?

Ultimately, the authors identified this regulatory relationship based on the interaction data. Is there any additional regulatory information that can be inferred given the network topology?

Reviewer #3 (Remarks to the Author):

In the manuscript "The Arabidopsis G protein network: linking G protein signaling, cell wall carbohydrates and plant morphogenesis," Klopflleisch et al. identify a G-protein interaction network and relate it to the possible role of heterotrimeric G proteins in cell wall composition. While the work represents the next logical step in the plant G protein signaling field and has generated an inventory of possible G protein-interacting proteins, it falls short on providing any possible mechanistic information on 10 plus years of plant G protein signaling research. There has been an incredible amount of data on G proteins' regulation of multiple plant hormone-signaling pathways, sugar sensing, control of cell division, etc.; however, no clear pathways seem to emerge from the interaction screen that would justify the role of G proteins in these pathways established by molecular-genetic studies. The interactomics data add one more set of responses possibly regulated by G proteins, the cell wall composition, without offering any possible explanation about how these proteins might be working. At this point, most of the data (except the interaction data itself) seems tangential.

Additionally, a number of concerns remain about the experiments per se. For example, a number of
proteins used for interaction screens are membrane-tethered or membrane-associated proteins. Such proteins are known to perform poorly in yeast 2 hybrid system. Obviously, many of the important interactions would have been missed using this system. Are there similar datasets available for G protein interactome in other organisms? If so, how do these sets of data compare? Moreover, the authors mention that they used three biochemically distinct forms of GPA1 for screenings. However, there is no data on whether it made any difference in the interaction partners identified.

It is interesting to note that the authors have identified the two known Ggamma proteins (AGG1 and AGG2) as interaction partners of AGB1; however, a third Ggamma protein that has been identified very recently (Chakravorty et al., Plant Journal) was not identified in their screens. It seems fairly odd given that the third protein (named AGG3) is a very strong interactor of AGB1. The data on stomatal density phenotype is not convincing. If this is part of the biologically relevant information the authors are going to provide for the interaction data they have generated, it needs to be done in a proper way. The stomata are not fully developed at the cotyledon stage and should be visualized on the fully developed, expanded leaves. The authors seem to replicate what is known in the literature, however for the new genes, having a single T-DNA insertion line and counting stomata on cotyledons is far from convincing.

Similar concerns remain for the experiments in which the authors have determined cell wall composition. First of all, there are morphological defects between the WT and gpa1 and agb1 mutants under normal growth and development conditions, in both roots and leaves. So detecting subtle changes in cell wall composition (and they are really quite subtle!) could be due to an inherent change in morphology, which has been explained in the past due to changes in rate of cell division. One could argue that subtle changes in cell wall composition contribute to the morphological defects, but at this stage it is not clear whether these changes are the cause or the effect. Even more perplexing is the use of liquid culture grown plants for such analysis. Plants grown in liquid culture for 28 days, under constant light and constant shaking, are bound to be strange and reveal biologically irrelevant information. As I mentioned earlier, the changes in cell wall composition are really minor and further experiments are needed to make any conclusions based on these data.

The correlation of gene expression data set with the G protein network does not add anything to the manuscript. It is again an example of indirect analysis to prove a point.

Overall, even though the data presented add to the information available about plant G proteins, it falls short of providing any further knowledge about how these proteins might work.

1st Revision - authors’ response 07 July 2011

Reviewer #1 (verbatim):

> The manuscript is very well written and addresses an important topic in plant systems analysis. Despite my overall favorable view of the manuscript, important questions need to be addressed.

1. For the Y2H analysis, the experimental section needs to expanded. Especially the well known artifact of autoactivation and spontaneous autoactivation need to be addressed. Additional experimental controls may need to be performed here.

RESPONSE: a) Every bait was tested for auto-activation even if that bait came into secondary screens as a prey. See page 4, line 26

b) Although labor intensive on our part, every positive interaction from the 200+ screens was re-cloned and tested again. Any interaction that did not survive this confirmation test was discarded from the list. This confirmation, which sets our interactome project apart from most others, was provided in the supplemental text and is now emphasized in the main text. See page 5, line 1

c) In addition to the Y2H field's list of common artificial interactions, we generated our own in-house list of artifacts on the basis of >2000 screens using more than 200 baits that we performed during the last decade (see Table in Supplemental Experimental Procedures). Prey clones representing any of these genes were removed from our interactome. This is now more clearly explained in the main text and the supplemental experimental procedures. See page 5, line 6
2. Validation: I applaud the authors for doing an experimental validation that is unbiased by previous knowledge (as computational analyses necessarily are) and to do so on a significant part of the network. However, the validation rate of 95% is extremely high and a reason for concern. It was previously shown using calibrated interaction assays (Braun et al., Nat Meth 2009), that different assays have distinct protein-interaction-detection profiles and under stringent conditions an orthogonal assay detects commonly detects <50% of interactions that were identified using a different assay.

RESPONSE: Braun et al. compared interactions found by different methods, but these interactions were completely unfiltered and therefore the expected lower (50%) confirmation rate is understandable. However, our approach, which was guided by Braun et al.’s recommendations, was far more selective and therefore the high confirmation rates that we observed are appropriate. First, as described above, all commonly known Y2H false-positives and genes by experience known to represent false-positives were discarded from further analysis. Second, all potential interactions revealed in each primary screen were re-tested in yeast, using re-cloned baits and preys to exclude false positives. Third, we applied a topological criterion for selection to be validated; each edge to be tested by BiFC must be part of a cluster. Specifically one of the two nodes in a cluster must have an interaction with at least two other proteins in the network. According to Braun et al., applying the latter two criteria strongly increase the confidence that a Y2H interaction is “real” and therefore these two criteria alone are expected to generate a much higher in planta confirmation rate. Given that we applied three stringent filters, it is not surprising that our validation was close to 95%. We describe this high stringency in the text. Page 6, line 14

> The concern here is that the BiFC assay is either done or scored with very promiscuous conditions and that the validation is therefore invalid. To address this concern, I recommend taking set of random proteins pairs, that er se cannot be expected to interact (no less than 20), and test this under the very same conditions, to obtain an estimate of the false positive rate of BiFC.

RESPONSE: The reviewer’s proposed experiment addresses the false positive rate of the BiFC technique in general. What is more important is providing a confidence level of our edges. This was tested many ways (more below). First, the BiFC vectors we used are second generation with a lower false positive rate (Waadt et al 2008 Plant J.; Uhrig et al 2007 Development). The false positive rate with this new reporter is lower than what most researchers are familiar with when the technique was first described. To assuage the reviewer’s concern, we added 24 additional negative interaction controls to the BiFC results. We tested another 24 nodes from the core interactome against two negative controls, one being a soluble protein (Arabidopsis thaliana histidine phosphotransfer protein AHP2) and the other a membrane protein (full-length membrane intrinsic protein PIP2a). That sums up to a total of 24 more negative controls added to our original four negative controls we showed in the previous version. We also added another 15 tests of edges of which 12 were validated, thus raising the total tests to 109. These are added to the supplement material. While this new data do not answer the question of „What is the false positive rate for the BiFC technique?“ they do support our high confidence level for BiFC validation.

3. I am also concerned about the fact that only pairs from the core-interactome are validated, whereas the manuscript describes a much larger network for a large part of which no estimation of the quality is provided. It is recommended to pick pairs for validation that either come to equal parts from the core and non-core dataset. Most appropriately interacting pairs should be picked at random from the 544 interactions and then both datasets will be represented to equal parts in the validation tests.

RESPONSE: The 109 tests (including controls) by BiFC took 2 full-time postdocs and 3 assistants an entire summer to complete. Obviously, this is a labor-intensive method and therefore, it is understandable why we would focus on the core interactome of nodes having at least two edges. However, for this revision, we added 15 more edges with eight being outside the core interactome. These new tests come from a functional cluster of calcium signaling elements. We feel that our work provides an unusually high rate of in vivo verification assays for a truly large-scale analysis.

> 4. The statement that ‘triangles are overrepresented in the network’ is meaningless and misleading and should be removed as this is a consequence of the biased experimental design and filtering. It could be replaced by ‘abundant’.
RESPONSE: We agree. The entire section was deleted. See page 5, line 27. Figure 1b panels illustrating these clusters was also deleted.

> 5. The numbers of co-expression correlated gene pairs in confusing. After some mulling of the sentence, I suspect that not all of the 544 pairs could be tested in all categories. However, this is not well explained and the ratio of correlated pairs versus tested pairs should be presented for each category. Further, why are 903 pairs tested when the interaction dataset only encompasses 544 protein pairs? This section needs to be completely reworked and more details on findings should be provided. It would be interesting, e.g. to see whether the pairs in a certain category have a higher co-expression correlation in appropriate datasets than pairs not in the category in the same transcription datasets.

RESPONSE: This section was completely reworked as the reviewer suggested. The original 903 set contained duplicates. These data were re-evaluated with duplications in the pairings and analysis removed. This showed a drastic reduction in the number of unique pairings analyzed and was closer to the 544 number mentioned elsewhere. Additionally, null permutations were performed on this dataset consisting of the unique pairs to identify the null distribution characterized by these data. The thresholds were similar to those produced by the gold standard data, which represented our training set. Therefore, we are confident in moving forward using the gold standard dataset as a training set since these unique pairs have true associations. These new data are now described in the main text (see page 6, line 8) and in the supplemental material.

> 6. For the cell wall findings an illustrative figure would be instructive. Potentially a cartoon of how different analyses fit together would help the reader. I am not entirely sure what is missing, but overall the figures accompanying this manuscript are less helpful than they could be. On the other hand, I am not entirely clear what the message of the network figures is and if it is necessary. Network representations can be very informative, however from the presented graphs I gather the message that the original bait proteins are most highly connected, which is hardly surprising and due to the experimental design. Perhaps the core network can be a panel of the last figure, and the cell wall synthesis proteins in the network could be highlighted to justify the authors investigation of this process.

RESPONSE: The biochemical and IR spectroscopic findings are insufficient to speculate extensively about direct or indirect pathways and interactions that could be portrayed in a cartoon. While small differences in monosaccharides proportions and in IR absorbance spectra are observed, they are statistically significant and typical for several cell wall mutants. These small differences often underlie larger changes in wall architecture and polymer composition at a cellular level, and this fine structure is revealed by the epitope labeling studies herein. However, we rewrote the text around this section to clarify the logic of the order of experiments, the limitations of each, and the conclusions.

Changing the order of the figures is not possible since the large amount of supplemental material that is related to figure 1 is presented by necessity in a defined order. The instructions to authors indicates that supplemental material must be related to a main figure and discussed soon after the main figure is cited. In addition, we feel that the present order reflects best the logic of the study and the order of the experiments.

> Reviewer #2 (verbatim):

>Overall, this manuscript is interesting. It presents a two hybrid analysis of proteins that interact with G proteins, initially in an unbiased manner, but at later points using a more biased approach. The authors do minor network analysis, and major in planta validation. They find an interesting novel interaction between G proteins and cell wall biosynthesis. Overall, the paper reads almost as two distinct studies - one with network analysis/coexpression - then jumps into a non-network analysis that does not use quantitatively derived network topology to characterize its biology.

Perhaps this could be addressed by some rewriting and some additional statistical analysis of the biological functions of the nodes within the network.
RESPONSE: Additional statistical analyses were performed. The GO categories (biological processes, cellular functions, molecular functions) were determined for the core network and compared to the genome-wide GO distribution. These new data were added to supplemental materials. We also analysed all genes for shared *cis*-regulatory elements and made this information available in the database using a “click and go” feature.

**1. Detected 544 interactions between 434 proteins - a comprehensive interactome.** How many times was a particular interactor found in each library for the first screen? The authors state "from the total of 206 unique prey proteins from the first round of screenings, 13 secondary baits were chosen because they interacted with two or more primary baits". What makes these baits unique? The fact that they were found once? If so - then I would say that this network is not comprehensive. There should be a threshold for the number of times a prey is found (sufficient coverage) to interact in any given library before deciding on the secondary screen. What if there are effectors that interact with just a single bait protein. Does this make them any less functionally important? Is there support for this in the literature?

RESPONSE: All the requested information (e.g. number of times found in the screen) is (and was) available in the database (http://bioinfolab.unl.edu/AGIdb). The username and password are provided in the Supplemental Information and here [user = gsignal password = rgs].

We would like to point out that the word 'unique' may be misleading - or could be misinterpreted. The number of times a prey is found is primarily proportional to the abundance of that clone in the library, not the strength or confidence of that interaction. Nonetheless, this information is still provided in the database. What might be more informative is the number of different libraries a prey was found with the same bait. Again, this could be an abundance issue. The data were also provided in the AtGIdb (Database).

Therefore, the key point is that every interaction we found in the screen was re-tested for validation. Many did not pass the re-test and therefore did not meet our stringent criteria and were therefore discarded. This was clarified further in the main text. See page 4 line 1 We do not conclude that potential interactions outside the core interactome are not interesting or any less important. Our emphasis on the core interactome is simply due to higher confidence for these interactions based on numerous filters and validations, including *in vivo* confirmation. Our goal was to provide a high quality network.

2. 45% of interactions were validated in planta This is a particularly impressive rate of validation.

RESPONSE: I think the reviewer meant to write that 45% (now more than 50% owing to additional BiFC assays performed for this revision) of the core interactions were tested, although it is true that most of these were also validated. With a focused interactome project, a high level of validation is possible (and perhaps should in future be expected – see also comments to points #1 and #2 by reviewer #1 above).

What is the relevance of the expression patterns of these signals? Is there evidence in the literature that these expression patterns generally reflect the *in vivo* subcellular location of these genes?

RESPONSE: Yes, these are translational fusions and therefore trafficking signals within bait and prey would direct them to their native plant compartments. If bait and prey are in the same compartment and do interact we would learn subcellular location information as well (Supplemental Figure S2).

*Transient transformation in tobacco seems a little artificial. More importantly, I would want to know if these gene pairs are at least expressed in the same context in at least one of the AtGenExpress arrays?*

RESPONSE: The reviewer’s concern is not clear to us – especially given the first sentence of point #3 below. Yes, we performed extensive analyses on co-expression of the encoding genes for edges in the entire interactome. Correlation coefficients and statistics were provided both in the supplemental material and in the database. Correlation (or not) of expression was performed using 4
different categories of gene expression profiles: development, mutant, anatomy, stimulus. This represents 4 different biological contexts which to our knowledge has never been done before. Since coexpression analysis is based on microarray data, correlation of coexpression is indicative of expression of the gene pairs throughout various tissues and conditions. Statistical analysis was possible because we trained our algorithm with a gold standard set of known interacting pairs. The most conservative threshold of 95% was applied and these gene pairs were indicated. Again, the conservative cut-off fit our goal for high quality data. We laid the foundation for others to mine the entire dataset deeply.

With regard to the statement about transient expression in tobacco we like to stress the fact that this procedure is common for high through put. Agrobacterium-mediated transient expression in Arabidopsis does not work reliably and is therefore impractical for larger test sets.

> 3. Co-expression analysis reveals that most are co-expressed (problems with null idstribution) I am concerned about the method of switching the pairs of the gold-standard proteins to create a null distribution. Do these conclusions hold when a set of genes are randomly chosen from the Arabidopsis genome? This particular null distribution isn't quite null since the gold standard positive set are also likely to be correlated in their expression. Their co-expression would be higher than that of the genome - they are likely being overly conservative, and aren't really testing if the co-expression levels are much higher than what would be expected by random chance. Perhaps even consider using the distribution of co-expression values from negatives obtained by the screening, and then compare the distribution of the positives from the screening?

RESPONSE: In the original manuscript, we analyzed the correlation of the gene pairs. The question is how do we determine if something is significant? We used the gold standard data to establish thresholds. This is unusual as normally there is not a secondary dataset with known underlying hypotheses, so we had an advantage. If we did not have the gold standard set, our null hypothesis would have to include the situation in which no true associations exist. Thus, to protect the amount of false positives in our result, the threshold would be set high. However using the gold standard data allows us to use a set that contains true associations in it, thus we can better define what true gene pairs look like when analyzed. The question is still, by looking at a set with true associations would you have a problem with setting the threshold too high?

> [further clarification of a point made above] 'Their co-expression would be higher than that of the genome - they are likely being overly conservative, and aren't really testing if the co-expression levels are much higher than what would be expected by random chance'

RESPONSE: For this, we looked back to the original gene pairs and computed their null distribution using the permutation approach. This is the normal approach for this type of analysis when a secondary dataset like the gold standards is unavailable. By doing this we want to make sure we have a similar null distribution to the gold standards, thus indicating we likely have true associations in the data. This is where further analysis was performed. The permutations were carried out on the gene pair data to compute null distributions. When computed, the thresholds came out slightly higher than the gold standard data (all within 0.03 so not too high). This indicates that the null distributions are quite similar and only slightly vary in the tail which supports us using data based on gold standards. Therefore the original analyses remain intact after additional computations were carried out.

> Also - the authors mention co-expression within network motifs within this section as justification for this analysis. However - no analysis was performed to see if there is any significant structure within this network, and any such analysis would be unjustified given the experimental design for the choice of additional baits in the second screen. I would therefore remove this sentence, and I would look at the structure of the network more quantitatively.

RESPONSE: We agree. We replaced our statistical statements with qualitative statements and we removed the sentences indicated. See page 5 line 27.

> [4a.] Reveal a novel role for G proteins in regulating cell wall modification. The authors mention protein clustering on page 7. I see no evidence of protein clustering. Is the presence of genes regulating the cell wall a significant observation? In a randomized network with the same structure,
what would the distribution be of genes with a cell wall annotation?

RESPONSE: Yes, at least it seemed so by eye. 49 of 434 proteins or 11% of the interactome nodes had a connection to cell walls whereas 6% of a random set of proteins in the genome were given this function by the same criteria. However, this is based on our current knowledge of cell walls and not on an unbiased GO annotation search. Therefore, we would not like to make claims about the abundance of cell wall-related proteins in the interactome in the manuscript. Instead, we set up the logic that we chose a line of investigation to establish proof of principle.

Actually, it does not matter if the interactome has a higher percentage of cell wall-related proteins or not. We would have followed the cell wall lead even if these cell wall genes did not seem over-represented. There had not been any previous connection between Arabidopsis G proteins and cell walls before and we felt this was an opportunity to test a novel hypothesis and make an important discovery.

Therefore, we rewrote this section to simply state that this was the line of inquiry that we chose. The reason, as discussed above, is that this is an area where a role of G proteins had not been discovered. See page 8 line 6 and 13.

> The cell wall analysis is really striking, particularly because, to the best of my knowledge, this is the first report linking G protein signaling to the cell wall. The biochemical analysis is well performed, as is the clustering of the antibody epitopes on the cell wall. The difference between the root and the shoot is also interesting. Are both these genes expressed equally in the root or the shoot, or are they expressed in different cell types in the shoot and root that would potentially explain these differences?

RESPONSE: It is not clear which two genes that the reviewer is asking about. To pursue this, we would need to take a gene by gene approach for cell specific expression (e.g. promoter::reporter, in situ hybridization, immunohistochemistries), which is clearly beyond the scope of the present work.

> Ultimately, the authors identified this regulatory relationship based on the interaction data. Is there any additional regulatory information that can be inferred given the network topology?

RESPONSE: We now linked all the genes encoding the nodes to AGRIS so that the user can quickly and effortlessly retrieve data on the cis-regulatory elements of the corresponding promoters. We analyzed two clusters that we call “stress clusters” because of the predominance of proteins known to play a role in stress adaptation. We found that most of the genes encoding nodes in these clusters contained cis-acting elements related to stress. New data are added to the database. See page 3 of the Supplemental Experimental Procedure section.

>Reviewer #3 (verbatim):

> In the manuscript "The Arabidopsis G protein network: linking G protein signaling, cell wall carbohydrates and plant morphogenesis," Klopffleisch et al. identify a G-protein interaction network and relate it to the possible role of heterotrimeric G proteins in cell wall composition. While the work represents the next logical step in the plant G protein signaling field and has generated an inventory of possible G protein-interacting proteins, it falls short on providing any possible mechanistic information on 10 plus years of plant G protein signaling research. There has been an incredible amount of data on G proteins' regulation of multiple plant hormone-signaling pathways, sugar sensing, control of cell division, etc.; however, no clear pathways seem to emerge from the interaction screen that would justify the role of G proteins in these pathways established by molecular-genetic studies. The interactomics data add one more set of responses possibly regulated by G proteins, the cell wall composition, without offering any possible explanation about how these proteins might be working. At this point, most of the data (except the interaction data itself) seems tangential.

RESPONSE: The power of a molecular systems approach is that it takes you into undiscovered areas. Thus, by nature, this discovery-driven approach does not extract mechanism but this approach definitely leads one to formulate a set of hypotheses on mechanisms that are tested over time. The interactome described here is leading us into many new areas not previously predicted by published mechanistic studies. We chose to pursue the possible cell wall connection. Specifically, the interactome prompted our hypothesis that xylan composition in the wall of G proteins is altered. Our
results from subsequent analyses and extensive cell wall epitope profiling were consistent with our hypothesis. The section that transitions from discovery through the screens and validation to the hypothesis testing was rewritten to highlight this feature of molecular systems biology. See page 8 line 13

> Additionally, a number of concerns remain about the experiments per se. For example, a number of proteins used for interaction screens are membrane-tethered or membrane-associated proteins. Such proteins are known to perform poorly in yeast 2 hybrid system.

**RESPONSE**: We did not find any preys that have transmembrane domains. The coding regions found for all preys are described in the database provided. We did not use any baits that were membrane-tethered or associated. For RGS1 (a 7-transmembrane protein), we used only the C-terminal cytoplasmic domain as bait. Myristoylation or prenylation sites on all baits were mutated to circumvent the problem that the reviewer noted. This information is now provided in the Experimental Protocols See page 1.

> Obviously, many of the important interactions would have been missed using this system.

**RESPONSE**: This, of course, could be true (as for any large-scale screen), but it is not obvious. To our knowledge, our interactome is based on the deepest set of screens ever performed. Nine divergent cDNA libraries (including one normalized) were interrogated. In addition, one core bait (the Gb subunit, AGB1) was screened 10 times in order to address the question of saturation; however a statistically supported conclusion about the level of saturation could not be reached and therefore we did not make such a conclusion. Admittedly, our Y2H screens might have missed some interacting membrane proteins (see also above). Labor-intensive proteomic profiling of G protein complexes should yield additional interactors that would never be found in Y2H screens (e.g. if phosphorylation is a requisite for interaction). These are for near future studies.

> Are there similar datasets available for G protein interactome in other organisms? If so, how do these sets of data compare?

**RESPONSE**: The reviewer is referring to what are called interologs. No, we did not find interologs in public eukaryotic interactomes. Given that Arabidopsis lacks all known canonical effectors described in animals (see Jones and Assmann 2004 EMBO Reports 5: 572-578), we were not surprised that interologs for our interactome did not exist. A statement about this was added. See page 6 line 4. Note that the lack of interologs might be in part due to the plant-specificity of some of the identified interactors (e.g. cell wall-related proteins).

> Moreover, the authors mention that they used three biochemically distinct forms of GPA1 for screenings. However, there is no data on whether it made any difference in the interaction partners identified.

**RESPONSE**: Yes, we did detect preys with the activated form of GPA1 that were not found by the inactive form as bait. The information is provided in the database (e.g. see VOZ1). However, this does not mean that these preys are form-specific effectors, so no conclusions were made. We invite the readers to test this possibility by extensive quantitative protein interaction methods. No change in text was made.

> It is interesting to note that the authors have identified the two known Ggamma proteins (AGG1 and AGG2) as interaction partners of AGB1; however, a third Ggamma protein that has been identified very recently (Chakravorty et al., Plant Journal) was not identified in their screens. It seems fairly odd given that the third protein (named AGG3) is a very strong interactor of AGB1.

**RESPONSE**: The reviewer is referring to the work of Chakravorty et al et al which was just published in the Plant Journal but this paper was not public information during the review process. However, we were aware of this work prior to publication because two seniors author of that work are co-authors on the present data set. First, Chakravorty et al. never detected Gg3 (AGG3) in Y2H screens, even though they screened extensively and that effort eventually resulted in the cloning of AGG1 and AGG2. Rather, they discovered AGG3 using bioinformatics and then they tested for interaction with Gb (AGB1). Second, AGG3 expression is restricted to guard cells and
floral/reproductive tissue and the actual levels of expression are quite modest; meaning that it would be difficult to fish it out even using a tissue specific library. Third, the reviewer is incorrect to conclude that the interaction strength between AGB1 and the full length AGG3 is very strong. It is true that there is a weak interaction between AGB1 and the full length AGG3 but it is much weaker than AGB1/AGG1 or AGB1/AGG2 (see Chakravorty et al Figure 2b for the proof). It is only when the long Cys-rich C-terminal domain was excised that the interaction became stronger than the AGG1/AGB1 or AGG2/AGB1 interaction. Thus, it was not likely that we would have found AGG3 as a prey in our extensive screens.

> The data on stomatal density phenotype is not convincing. If this is part of the biologically relevant information the authors are going to provide for the interaction data they have generated, it needs to be done in a proper way. The stomata are not fully developed at the cotyledon stage and should be visualized on the fully developed, expanded leaves. The authors seem to replicate what is known in the literature, however for the new genes, having a single T-DNA insertion line and counting stomata on cotyledons is far from convincing.

RESPONSE: The point of testing stomatal density (the precise term is stomatal index, SI) was to test the hypothesis that the cell wall proteins in the interactome shared a connection to development as do the G proteins. We chose a G protein developmental phenotype that was easily scored and quantified and suitable for a systems biology approach (i.e. we chose SI). Therefore, we followed exactly the protocol originally used to measure SI (Zhang et al., Dev. Biol. 2008) for the Arabidopsis G protein mutants. This protocol calls for the use of cotyledons of young seedlings, not fully-expanded leaves. However, there is a publication describing the SI of fully developed leaves (Nilson and Assmann, Plant Physiol. 2010), but the point of that work was to relate the stomatal density to leaf conductance, not development. Moreover, the overall conclusions regarding the SI phenotype of cotyledons vs. fully-expanded leaves of the G protein mutants are the same, so this argument becomes a moot point here. A full developmental (e.g. time course) characterization could be done but at this juncture, with little justification.

In some cases, only a single mutant allele was available to test. We feel that a more comprehensive analysis (i.e. testing more alleles, time course etc.) lies beyond the scope of this succinct Report.

> Similar concerns remain for the experiments in which the authors have determined cell wall composition. First of all, there are morphological defects between the WT and gpa1 and agb1 mutants under normal growth and development conditions, in both roots and leaves. So detecting subtle changes in cell wall composition (and they are really quite subtle!) could be due to an inherent change in morphology, which has been explained in the past due to changes in rate of cell division. One could argue that subtle changes in cell wall composition contribute to the morphological defects, but at this stage it is not clear whether these changes are the cause or the effect. Even more perplexing is the use of liquid culture grown plants for such analysis. Plants grown in liquid culture for 28 days, under constant light and constant shaking, are bound to be strange and reveal biologically irrelevant information. As I mentioned earlier, the changes in cell wall composition are really minor and further experiments are needed to make any conclusions based on these data.

RESPONSE: The cell wall differences are detectable and real. One will not find many published mutations in cell wall genes that give profound cell wall composition differences. The few that do, confer gross developmental phenotypes and there is where one gets to the „chicken or egg” problem. One of the advantages of using Arabidopsis to study G protein signaling is that complete genetic ablation of the G protein complex does not confer profound developmental phenotypes, thus permitting conclusions to be made about signaling (rather than aberrant development indirectly affecting signaling). The advantage of a systems approach is that you are led to subtle phenotypes that often would not be found otherwise.

The xylan differences we reported were obtained using both standard plate-grown seedlings and seedlings grown in liquid medium; we reached the same conclusion for both. This is described in the supplemental Experimental Procedures. Specifically, the original biochemical analyses of cell walls that lead to the final cell wall epitope profiling (liquid-culture grown seedlings) were obtained using 7-day-old seedlings grown on standard plates under short day light regimes. (information described in supplemental) These analyses pointed to statistically-significant differences in xylan composition
that then prompted the extensive cell wall epitope profiling experiments.

> The correlation of gene expression data set with the G protein network does not add anything to the manuscript. It is again an example of indirect analysis to prove a point.

RESPONSE: We agree that the correlation of gene expression is indirect, but it is yet another of several means that we utilized to validate the interactome. This information is now assembled into a database suitable to mine deeply for future studies.

> Overall, even though the data presented add to the information available about plant G proteins, it falls short of providing any further knowledge about how these proteins might work.

RESPONSE: Mechanisms for the hundreds of interactions that we reported are for future in-depth studies. The systems approach led us to a previously unknown role for G proteins in cell wall architecture and composition. Any further studies are beyond the scope of the present work.

2nd Editorial Decision 05 August 2011

Thank you again for submitting your work to Molecular Systems Biology. We have now heard back from the three referees who agreed to evaluate your revised study. As you will see, the referees generally agreed that the revisions had significantly improved this work, and the first two reviewers are now largely supportive. There appear to be, however, some remaining important concerns, which we would ask you to carefully address in a final revision of the present work.

The third reviewer is the least supportive, and seems to have concerns regarding the conclusiveness of the cell wall composition analyses. As emphasized in our previous decision letter, the editor feels that extensive additional experimentation analyzing additional insertion lines lies outside the scope of this short Report, especially considering the positive comments from the other two reviewers. Nonetheless, since these experiments provide the main evidence supporting the functional relevance of the G-protein interaction network, it is essential that the existing data is conclusive and rigorous. In this regard, and considering the third reviewer's concerns, the editor feels that additional information is needed to conclusively support the significance and reproducibility of the ELISA-based glycome profiles. Could you please address/clarify the following points in detail?

-- Please provide the full data for the glycome profiling experiments as supplementary material, including individual measurements for replicate experiments. Molecular Systems Biology generally requires that all relevant experimental data is fully released with published works.

-- The statistical methods used to calculate the significance of the glycan epitope abundance differences was not clear (changes in xylan-2 and -3 levels are described as significant twice on page 9). Please describe this in your Methods section, and include p-values in the results section whenever significance is claimed. The editor notes that there is a multiple test correction issue here, and methods used to correct p-values for multiple testing should be clearly described in the Methods.

-- Please clarify what work is referred to by "Pattathil et al. unpublished". Is this a reference to the methods described in Pattathil et al. 2010? Or to another unpublished work? If there is another related work in preparation please provide a draft with your revision.

-- Please include some brief additional discussion of the results the last reviewer found strange (i.e. inconsistent mole % changes).

In addition, please address the other more minor issues raised by the other two reviewers. Importantly, the second reviewer, while generally quite positive, also has some remaining doubts about the evidence connecting the interaction network to cell wall biology, and feels that some more rigorous statistical testing (e.g. point #5) would improve this work.

Thank you for submitting this paper to Molecular Systems Biology.
Yours sincerely,

Editor - Molecular Systems Biology
msb@embo.org

----------------------------------------

REFEREE REPORTS
----------------------------------------

Reviewer #1 (Remarks to the Author):

Review of second submission of Manuscript Number: MSB-11-2927

Title: Arabidopsis G-protein Interactome
Klopffleisch et al.

The authors have very well addressed the concerns raised by the reviewers and much improved the manuscript. At this stage, all questions of the reviewers have satisfactorily been addressed in the response letter. Below are few requests to clarify the manuscript and to translate some of the clarity of the answers in the response letter into the main text of the manuscript.

1. High validation rate of BiFC.

The author's response states clearly that both the experimental and topological filters applied to the verified interactions contribute to the high validation rate. This should be more explicitly stated in the main text. I am suggestion 2 additions:

Pg.6 ln 10: replace "three stringent filters" with "(three) stringent experimental and topological filters"

In addition, it is still important to mention the limitations of the negative controls. The authors have made a good and good-faith effort to determine the FP rate of the BiFC and the fact that a second generation split-YFP was used is likely to contribute to the good performance of the assay. At the same time, it is also likely, that the background of the assay is inadequately described by the experiments using a limited set of proteins. In light of other journals recommending benchmarking against available reference sets for various organisms, it is important to acknowledge the state-of-the-art with respect to benchmarking validation assays. I am recommending including a brief discussion along the following lines in the main text:

"Our validation rate is higher than expected given the previously observed detection overlap between Y2H and orthogonal validation assays [ref]. This is likely due to the applied filtering of validated interactions, although formally it cannot be excluded that use of a larger and more diverse set of random negative controls would detect a higher background and hence necessitate more stringent scoring."

Lastly, as the authors will undoubtedly be aware large-scale interactome map of Arabidopsis was recently published in Science. It cannot be expected that the authors do a detailed comparison of their data to this new data set this late in the process, but the discussion should include a mention of this related data set and claims may need to be adjusted.

Reviewer #2 (Remarks to the Author):

The manuscript is recommended for publication as long as the following concerns are addressed.

1. These are two distinct studies - one with network analysis/coexpression - then jumps into a non-network anlaysis that does not use quantitatively derived network topology to characterize its
My main concern was that there was no indication of how the network topology related to the findings of proteins associated with the cell wall. Ultimately, the author's rationale is justified - the proteins associated with cell wall synthesis were identified with the high-confidence interactome, and clear phenotypes were identified. Biologically, this is extremely exciting as cell wall synthesis is a biological process for which phenotypes are not easily found because of extreme redundancy. Overall, their inclusion to address this issue is good and largely addresses my concern. Despite that, there are some statistical concerns with their analysis that can be easily addressed:

- GO category analysis - how is statistical significance defined? The authors present a ratio, but it is unclear how statistically significant this is. Is this supposed to be determined by the raw score? What is the raw score? A Fisher's exact test (significance using hypergeometric distribution) is appropriate here. Also, what does "ratio B/A" mean?
- cis-regulatory analysis - it is unclear how these "clusters" were defined. Please define in the manuscript text.

2. The rewritten manuscript makes the description of interactions and their confidence much more clear. The point that all manuscripts were retested by independent cloning and re-testing is an extremely important one - and a massive amount of work undertaken by the authors. This work will no doubt have implications on other interactome studies - their level of filtering certainly decreases the false negative ratio.

3. The number of interactions biologically validated is very impressive, and I believe, amongst the highest in any type of study like this. It is worth mentioning however, that these are transient assays and still just suggest that these interactions occur in planta.

4. Coexpression analysis - this is much more clear, and the authors state that they revisited their approach by looking at the null distributions in the absence of a gold standard dataset. Some discussion should be made however about the cases where co-expression appears to be much higher in development (>0.905), and even though the correlation is lower for stimulus (0.138) and mutation (0.135) with near 50% of gene pairs showing significant correlation of gene expression. The assumption here is that the gene pairs identified represent as of yet uncharacterized biological processes as determined by gene pairs being co-expressed.

Also, this statement is unclear: "Overall, the average expression correlation was significantly higher than the reliability of the established G protein interaction network". Which established G protein network do they mean? What do they mean by reliability? Please be more concise and clear.

5. With respect to the connection with the cell walls - determining whether or not this GO category annotation ratio is significant - again - using a Fisher's exact test would provide a useful quantitative measure to introduce why this was studied.

6. Computational mining of available cell type expression data (AtGenExpress or EFP browser) of genes functionally determined to be associated with the cell wall would at least give a clue as to where these genes are generally deployed to regulate cell wall synthesis or if they are cell type-specific.

Reviewer #3 (Remarks to the Author):

In the revised version of the manuscript "The Arabidopsis G protein network: linking G protein signaling, cell wall carbohydrates and plant morphogenesis," Klopffleisch et al. have addressed some of the concerns raised in the previous version of the manuscript and have added some additional information. The manuscript is focused on the identification of a G-protein interaction network. In this version, the authors have added additional information on the controls that were used to test the interactions and on the nature of some of the bait proteins i.e. only the soluble portion of RGS1 was used and the potential prenylation motifs in proteins were mutated to avoid membrane targeting. The authors do not identify any interlogs in other species which they attribute to the unique nature of plant G protein signaling. This is still a bit strange as most of the 'intra G-protein complex interactions" are conserved across species (between Galpha and Gbeta, Gbeta and
While the interaction data per se are solid, the biological interpretation of the data is not convincing. For example, a change in mole % of xylose in agb1 and gpa1agb1 double mutants compared to the wild type plants is one of the important, biologically relevant findings of this report. So how do the authors justify a decrease in xylose level due to lack of AGB1, no effect due to lack of GPA1, but a significant increase in xylose level when both GPA1 and AGB1 are missing? Similarly, use of single T DNA insertion lines to report a subtle phenotype in stomatal indices is not convincing. The new data reported on GO analysis does not add anything to the manuscript. Similarly, a stress induced gene cluster that includes known ABA/stress responsive proteins will have an over representation of ABRE-like elements or CACTG motif in the promoters. This is not evidence that why it "further supports the notion that these genes are involved in ABA-and/or dehydration responses".

2nd Revision - authors’ response 15 August 2011

Editor's Requests

-- Please provide the full data for the glycome profiling experiments as supplementary material, including individual measurements for replicate experiments. Molecular Systems Biology generally requires that all relevant experimental data is fully released with published works.

RESPONSE: Raw data now provided in supplemental on-line materials. See Table S1.

-- The statistical methods used to calculate the significance of the glycan epitope abundance differences was not clear (changes in xylan-2 and -3 levels are described as significant twice on page 9). Please describe this in your Methods section, and include p-values in the results section whenever significance is claimed. The editor notes that there is a multiple test correction issue here, and methods used to correct p-values for multiple testing should be clearly described in the Methods.

RESPONSE: The confidence levels were added. See page 10, line 11. The statistical analyses were described in the Experimental section. See page 10

-- Please clarify what work is referred to by "Pattathil et al. unpublished". Is this a reference to the methods described in Pattathil et al. 2010? Or to another unpublished work? If there is another related work in preparation please provide a draft with your revision.

RESPONSE: This was deleted. This was an oversight; the published work cited at the end of this sentence is the correct one. See page 9, line 29.

-- Please include some brief additional discussion of the results the last reviewer found strange (i.e. inconsistent mole % changes).

RESPONSE: Opposite phenotypes are signatures in G signaling and evoke specific conclusions. The opposite phenotype indicate that the Ga subunit, not the Gbg dimer, is the operative element. This was clarified with an additional sentence See page 9, line 19.

When revising your work please reformat this work according to Molecular Systems Biology guidelines. Please check our citation format requirements, and please include a Materials and Methods section in the main manuscript, incorporating some or all of the supplementary experimental procedures.

RESPONSE: 1) Citation formatted correctly. 2) Author contributions added. 3) The experimental section was left as supplemental material because this is over 10 pages long and carries 37 unique references. The large number of techniques and the extensive statistical analyses utilized required extensive explanation. 4) text on the stress cluster was deleted (reviewer #3 agrees). See page 7, line 3. This new information (full cis element analysis of all the core interactome genes) is available to readers from the associated database. New tools were also added.
Reviewer #1

Pg.6 In 10: replace "three stringent filters" with "(three) stringent experimental and topological filters"

RESPONSE: Replaced as suggested. See page 6, line 16.

Include: “Our validation rate is higher than expected given the previously observed detection overlap between Y2H and orthogonal validation assays [ref]. This is likely due to the applied filtering of validated interactions, although formally it cannot be excluded that use of a larger and more diverse set of random negative controls would detect a higher background and hence necessitate more stringent scoring.”

RESPONSE: Sentence added as suggested. See page 6, line 17.

Lastly, as the authors will undoubtedly be aware large-scale interactome map of Arabidopsis was recently published in Science. It cannot be expected that the authors do a detailed comparison of their data to this new data set this late in the process, but the discussion should include a mention of this related data set and claims may need to be adjusted.

RESPONSE: This work was cited and claims adjusted. Two Arabidopsis interactome papers appeared as articles in Science last week (July 29). One was an 8000 x 8000 test with no associated biology. The second paper was like our work under consideration; it was a focused interactome with biology on disease resistance. See page 12, line 8.

Reviewer #2

“... there are some statistical concerns with their analysis that can be easily addressed:
- GO category analysis - how is statistical significance defined? The authors present a ratio, but it is unclear how statistically significant this is. Is this supposed to be determined by the raw score? What is the raw score? A Fisher’s exact test (significance using hypergeometric distribution) is appropriate here. Also, what does "ratio B/A" mean? - cis-regulatory analysis - it is unclear how these "clusters" were defined. Please define in the manuscript text.”

RESPONSE: The Fisher exact test was applied and the confidence level now provided. The use of B/A ratio was dropped. A better explanation of how the values were obtained is now provided. See page 17, line 4.

Coexpression analysis - this is much more clear, and the authors state that they revisited their approach by looking at the null distributions in the absence of a gold standard dataset. Some discussion should be made however about the cases where co-expression appears to be much higher in development (> .905), and even though the correlation is lower for stimulus (.138) and mutation(.135) with near 50% of gene pairs showing significant correlation of gene expression. The assumption here is that the gene pairs identified represent as of yet uncharacterized biological processes as determined by gene pairs being co-expressed

RESPONSE: Discussion of these was added to the Experimental Protocol Supplemental material. see page 6.

“...this statement is unclear: "Overall, the average expression correlation was significantly higher than the reliability of the established G protein interaction network". Which established G protein network do they mean? What do they mean by reliability? Please be more concise and clear. “

RESPONSE: See page 6.

“With respect to the connection with the cell walls - determining whether or not this GO category annotation ratio is significant - again - using a Fisher's exact test would provide a useful
quantitative measure to introduce why this was studied. “

RESPONSE: A sentence was added that despite a lack of statistical support by Fisher Exact test, we nonetheless notice a lot of cell wall-related proteins in the interactome. See page 8, line 17

Computational mining of available cell type expression data (AtGenExpress or EFP browser) of genes functionally determined to be associated with the cell wall would at least give a clue as to where these genes are generally deployed to regulate cell wall synthesis or if they are cell type-specific.

RESPONSE: No change. No time or space left for another round of analyses. With so many good ideas coming from the reviewers, it is clear that the community will, no doubt, find a lot to analyze once the data becomes public.

Reviewer #3.

So how do the authors justify a decrease in xylose level due to lack of AGB1, no effect due to lack of GPA1, but a significant increase in xylose level when both GPA1 and AGB1 are missing?

RESPONSE: There is confusion; we did not show gpa1 phenotypic data nor did we conclude: “no effect due to lack of GPA1”. Our response was already discussed in the editor section (above). Basically, the opposite phenotypes for Gb alone vs Ga and Gb together is a classic signature for Ga being the primary signaling element. A sentence was added to clarify. See page 9, line 19.

... a stress induced gene cluster that includes known ABA/stress responsive proteins will have an over representation of ABRE-like elements or CACTG motif in the promoters. This is not evidence that why it "further supports the notion that these genes are involved in ABA-and/or dehydration responses

RESPONSE: We agree and took the opportunity to slim down the manuscript by deleting this paragraph. See page 8, line 3.

3rd Editorial Decision 15 August 2011

Thank you again for submitting your work to Molecular Systems Biology. I have now had time consider your revisions, and I agree that the changes made have sufficiently satisfied the reviewers remaining concerns. This work is likely to be appropriate for publication in Molecular Systems Biology after some final format and data issues are addressed in a last minor revision.

1. Please provide the entire interaction dataset as a supplementary file. Molecular Systems Biology generally requires that all key data associated with a manuscript are included as supplementary material to ensure their longevity and accessibility. In addition, it was not clear whether these interactions had been deposited in AtPID (this database was not working when I tested it today). MSB generally requires that authors deposit interaction in a public database, and encourages authors to submit to repositories that are members of the IMEX consortium (http://www.imexconsortium.org/). The Methods section of the main manuscript should refer readers to these supplementary resources and mention any accession numbers that accompany submission to public databases.

2. Please see the attached manuscript, where I have made a series of suggestions to help bring this work in line with our Report format. In general, this involves moving some sections of the text to a new Materials and Methods section, and condensing other portions of the Results. Remaining methods not included in the main manuscript can be retained in the Supplementary Information pdf. This pdf should also include the Supp. Figures (except for Supp. Fig. 1), and should begin with a Table of Contents that describes all supplementary material including with this manuscript. With these changes the Title page through Conclusion runs about 22,000 characters, which is acceptable. The reference section is a bit long, but that seems warranted for this work (nonetheless, please check and remove any unnecessary citations).
When submitting your revision, please provide
-- the manuscript text in LaTeX, RTF or MS Word format
-- a letter with a detailed description of the changes made in response to the referees. Please specify clearly the exact places in the text (pages and paragraphs) where each change has been made in response to each specific comment given
-- three to four 'bullet points' highlighting the main findings of your study
-- a 'standfirst text' summarizing in two sentences the study
-- a "thumbnail image" (width=211 x height=157 pixels, jpeg format), which can be used to highlight your paper on our homepage.

The order of the authors in the final publication is determined solely by the author list on the manuscript title page, so don't worry about the order in the manuscript tracking system. Please declare the excel tables as "datasets" when uploading to avoid pdf conversion (regardless the system always retains the original uploaded file, so we can undo pdf conversions on our end).

Please feel free to contact us if you have any questions regarding these points.

Thank you for submitting this paper to Molecular Systems Biology.

Yours sincerely,