molecular systems biology

Epigenetic epistatic interactions constrain the evolution of gene expression

Solip Park and Ben Lehner

Corresponding author: Ben Lehner, EMBL-CRG Systems Biology Unit

Review timeline:

Submission date:	26 September 2012
Editorial Decision:	5 November 2012
Revision received:	17 December 2012
Acceptance letter:	7 January 2013

Editor: Thomas Lemberger

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 5 November 2012

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 manuscript. As you will see from the reports below, the referees find the topic of your study of potential interest. They raise, however, several concerns which should be convincingly addressed in a revision of this work. The constructive suggestions provided by the reviewers are very clear in this regard and refer to additional analyses and controls to better support the major conclusions you wish to draw.

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.

Referee reports:

Reviewer #1 (Remarks to the Author):

Summary:

The authors report a number of new and interesting results relating genetic interaction degree to variability in gene expression. The authors propose that a high number of genetic interaction partners may place selective pressure on the gene's expression, requiring low variability. The authors are thorough in their analysis by considering three different types of expression variability (noise, plasticity, evolution), they cover evidence of mechanisms of gene regulation (promoter

architecture), and suggest a possible cause (gene duplication) for a set of high-degree genes that also have high variability.

The relationship between gene expression and genome-wide distribution of genetic interactions has not previously explored. As gene expression regulation likely has a large effect on the phenotypes of individuals, this is a concept that will be very important in understanding (epi)genetic interactions in humans and how these vary throughout populations. I expect these findings will be of broad interest. The manuscript could be strengthened with attention to a few relatively minor issues and more in-depth exploration on a few fronts.

Major comments:
(1) The authors report correlations on non-binned data in the text, but the figures report correlation on means obtained from binning into groups of 100 genes (e.g. Figs 2-5). The trends and the conclusions the authors derive from them are solid, but this discrepancy in how the results are presented is a bit confusing and potentially misleading about the strength of the reported effects. I would suggest that the authors (a) not report correlation after binning, but instead include the global correlation at the gene level on the figure itself or in the figure legends, and (b) change the axes labels of the figures (or the style of the figures) so they clearly indicate that means of binned genes are being plot (e.g. "Mean number of genetic interactions").

(2) What is the role of genes' expression level in the reported trends? For example, if one controls for expression level, are the correlations between interaction degree and the various kinds of noise still significant? One possibility is that the number of genetic interactions per gene is really positively correlated to expression level and that there is simply a negative correlation between expression level and expression noise. (the authors have already done this analysis with fitness defect, but a similar analysis with expression level would be worthwhile)

(3) For the human gene divergence analysis, does the negative correlation between GI degree and expression divergence persist after controlling for fitness defect or expression level?

(4) The enrichment for duplicate genes among the highly variable genetic interaction hubs is very interesting. A few follow-up questions that would be interesting to explore: What is the status of the duplicate partner(s) of these genes? Do they also tend to be a hub, or do these pairs fit the frequently observed "asymmetric" class (VanderSluis et al. MSB 2010). What is the expression variability of the duplicate partner(s)? Are there differences in their mean expression level? Do they share overlapping genetic interactions? Related clarification question: what are the degree cutoffs in Supplementary Fig. 3 (labeled 3, 5, and 10)? Are these percentiles, actual # of interactions? I assume these must be percentiles since the median number of interactions is 22.

(5) Related to the duplicate comment above, it may be informative to explore the relationship between hubs and their interaction partners in general. For each hub gene, are its interactors similar (among themselves or compared to the hub gene) in their expression variability? What about their expression levels? Are these trends different for highly variable hubs vs. less variable hubs? Is there any relationship with co-expression (across conditions) for variable vs. non-variable hubs and their interaction partners? These points would not be necessary to address in a revision, but might be interesting to consider.

Minor comments:
(1) For this sentence "Indeed we find that genetic interaction hubs with variable expression are twice as likely to have gene duplicates as other genetic hubs, and in total have more than six times as many duplicates (Figure 6 and Supplementary Figure 3).", the "in total have more than six times as many duplicates" statement is confusing to me-this seems to imply that the class of variable expression hubs is much larger than the non-variable expression hubs class.

Reviewer #2 (Remarks to the Author):

In this manuscript, Park and Lehner examine the association between genetic interactions and different gene features in order to examine whether genetic interactions could constrain expression variation. The idea is simple and very clever. If Gene A and Gene B show a genetic interaction and their expression is noisy, cells in which gene product A and gene product B are in low abundance will suffer from this genetic interaction, even in wild-type cells. The authors find that in yeast genetic interaction data support their model: genes that have a large number of genetic interactions show less variation in expression levels (noise, plasticity and evolutionary divergence). The model is very interesting (although presented in previous papers by the same author so not completely novel). There are several limitations to the study however.

Major comments:

-it seems that positive and negative genetic interactions should lead to different predictions but these are not considered here. The majority of genetic interactions are negative but the two types should be analyzed separately as to better support the model.

-all of the variables that the authors correlate with the number of genetic interactions are also correlated with each other, e.g. noise, plasticity, diversity within species and divergence between species. Thus, it is not clear if all of the analyses support different predictions or are in fact measuring the same thing every time. This is not satisfying and should be dealt with in an appropriate multivariate analysis.

-the authors seem to imply that there is a causal relationship between the number of genetic interactions and a reduction of expression variation, i.e. genetic interactions constrain the evolution of and regulation of gene expression. The authors do not discuss the possibility that both could be caused by one or several factors that act to constraint or promote these two variables such that there is no causal relationship between the two. For instance, genes with a large number of genetic interactions are involved in a large number of biological processes, which requires that they are stably expressed. Other hypothesis like this could be put forward and test to find the most likely connections.

-overall, there is a lack of mechanistic explanation for the trends observed. The yeast data is rich and it should be easily feasible to find examples from the data where a mechanistic basis could be detailed. The manuscript is largely abstract (genetic interactions are indirect measurements of associations between pathways or genes and do not always translate directly into clear mechanisms) and lacks mechanistic explanations.

-the correlation coefficients reported in the text are often not the same as those reported in the figures. Correlation coefficients should not be calculated on binned data although binned data can be used for data representation.

Reviewer #3 (Remarks to the Author):

The main finding reported in this paper is that the number of genetic interaction a gene has correlates with the gene expression variability on three time scales: responsiveness to changing conditions, stochastic variations between genetically identical individuals and evolutionary divergence.
The authors provide the interesting interpretation to this result, suggesting that reduced expression variability reflects the possible deleterious effect of co-reducing the expression of synthetic-lethal partners, which could limit growth and thereby be selected against. For genes that have multiple synthetic interacting partners, modulating expression might be particularly problematic in this regards.

I find this result interesting and worth publication. Yet, I'm not sure it is fully established. Several controls should be considered:

1. The authors relate only to the number of gene interacting partners. What would be the prediction for interacting pairs? Are they expected to be inversely correlated in expression to reduce the possibility that they will both be low at the same time? Since hubs in the genetic interaction network are characterized by multiple properties (many of which are connected) it will be highly useful to try and support the hypothesis based on distinct (and perhaps less-connected) synthetic interaction pairs.
2. In this context, would the model predict distinct behavior of positive vs. negative interacting partners? This would provide an additional support for their model.
3. Can the authors suggest how to experimentally test their model? Can they support the notion that (moderate?) down-regulation of synthetic interacting partners reduces growth rate? Do they see evidence for this in the budding yeast literature? How would it fit with the idea that cells are largely robust to quantitative variations in protein levels?
4. My understanding is that the analysis is performed on a single dataset only. Although quite large, it is far from being comprehensive and there could be some issues relatd to the particular genes examined there. There are now many additional datasets available, in particular from the labs of Jonathan Weissman and Kevin Kogan (in addition to the datasets coming from Toronto). The authors should verify that the results extend also to those different datasets.
5. It will be nice to discuss the correlates of the synthetic interaction dataset used, as described in the original publication. In particular, the differences between the previously observed correlates (e.g. expression levels, phenotypic capacitance) and the present result should be emphasized. Those previous results might be helpful for the model interpreting the present data as well.
6. Correlation strength: In the beginning of the papers, the authors report very low correlation values (<0.2) while at the second half those correlation increase dramatically. Examining the method section, we believe this difference reflects the analysis method: first value corresponds to 'bare' correlation, while the second results from correlating the averaged windows. Now this averaging can artificially increase the correlation values. If this analysis is correct, the authors should be explicit about how they measure correlations and report the real values rather than artificially inflated ones.

1st Revision - authors' response 17 December 2012

Response to Reviewers' Comments

Reviewer #1

The authors report a number of new and interesting results relating genetic interaction degree to variability in gene expression. The authors propose that a high number of genetic interaction partners may place selective pressure on the gene's expression, requiring low variability. The authors are thorough in their analysis by considering three different types of expression variability (noise, plasticity, evolution), they cover evidence of mechanisms of gene regulation (promoter architecture), and suggest a possible cause (gene duplication) for a set of high-degree genes that also have high variability.
The relationship between gene expression and genome-wide distribution of genetic interactions has not previously explored. As gene expression regulation likely has a large effect on the phenotypes of individuals, this is a concept that will be very important in understanding (epi)genetic interactions in humans and how these vary throughout populations. I expect these findings will be of broad interest. The manuscript could be strengthened with attention to a few relatively minor issues and more in-depth exploration on a few fronts.

We are grateful to the reviewer for their positive assessment of our work. Here, we present our responses to the reviewer's comments and updates to the manuscript, which we hope the reviewer will find satisfactory.

Major comments:
1. The authors report correlations on non-binned data in the text, but the figures report correlation on means obtained from binning into groups of 100 genes (e.g. Figs 2-5). The trends and the

conclusions the authors derive from them are solid, but this discrepancy in how the results are presented is a bit confusing and potentially misleading about the strength of the reported effects. I would suggest that the authors (a) not report correlation after binning, but instead include the global correlation at the gene level on the figure itself or in the figure legends, and (b) change the axes labels of the figures (or the style of the figures) so they clearly indicate that means of binned genes are being plot (e.g. "Mean number of genetic interactions").

We agree and now only report correlations for the non-binned data. We have also changed the figure axes labels to 'mean number of genetic interactions'.

2. What is the role of genes' expression level in the reported trends? For example, if one controls for expression level, are the correlations between interaction degree and the various kinds of noise still significant? One possibility is that the number of genetic interactions per gene is really positively correlated to expression level and that there is simply a negative correlation between expression level and expression noise. (the authors have already done this analysis with fitness defect, but a similar analysis with expression level would be worthwhile)

As shown in Supplementary Figure 3B in the revised manuscript, the negative correlation between the different measures of expression variation and genetic interaction degree remains when controlling for expression level.

3. For the human gene divergence analysis, does the negative correlation between GI degree and expression divergence persist after controlling for fitness defect or expression level?

We confirmed the negative correlation between genetic interaction degree and expression divergence after controlling for fitness defect or expression level (Revised Supplementary Figure 5).

4. The enrichment for duplicate genes among the highly variable genetic interaction hubs is very interesting. A few follow-up questions that would be interesting to explore: What is the status of the duplicate partner(s) of these genes? Do they also tend to be a hub, or do these pairs fit the frequently observed "asymmetric" class (VanderSluis et al. MSB 2010). What is the expression variability of the duplicate partner(s)? Are there differences in their mean expression level? Do they share overlapping genetic interactions?

In the revised manuscript we have further explored the properties of the duplicates of genes with many genetic interactions (Revised Supplementary Figures 4A - D). To summarize: i) duplicate partners of both variable and non-variable genes with > 25 genetic interaction partners are likely to have more genetic interaction partners than duplicate partners of genes with ≤ 25 interactions (Supplementary Figure 4A), ii) duplicate partners of highly (or lowly) variable genetic hubs are also likely to be highly (or lowly) variable (Supplementary Figure 4B), iii) duplicate partners of both variable and non-variable genetic hubs are more likely to be highly expressed than duplicate partners of non hubs (Supplementary Figure 4C), and iv) duplicate partners of both variable and non-variable genetic hubs are more likely to have shared genetic interactions than duplicate partners of non hubs (Supplementary Figure 4D).

Related clarification question: what are the degree cutoffs in Supplementary Fig. 3 (labeled 3, 5, and 10)? Are these percentiles, actual # of interactions? I assume these must be percentiles since the median number of interactions is 22.

The degree cutoffs indicated the actual number of genetic interactions. We only considered genes that are highly or lowly variable in all of the types of gene expression variation (i.e. noise, responsiveness, mutation variation, *trans* variability, expression divergence and inter-strain variation), which is a very restricted group and so we used a low cutoff to split the genes. Prompted by the reviewer's comment we have now modified the definition of highly connected gene to genes with degree at least 25 (Revised Figure 6).

5. Related to the duplicate comment above, it may be informative to explore the relationship between hubs and their interaction partners in general. For each hub gene, are its interactors similar (among themselves or compared to the hub gene) in their expression variability? What about their expression levels? Are these trends different for highly variable hubs vs. less variable hubs? Is

there any relationship with co-expression (across conditions) for variable vs. non-variable hubs and their interaction partners? These points would not be necessary to address in a revision, but might be interesting to consider.

This is an interesting suggestion. We analyzed the relationship between highly connected genes (genes with degree at least 25) and their genetic interaction partners. To summarize: i) genetic interaction partners of highly (or lowly) variable genes are likely to be highly (or lowly) variable (Reviewers only Figure 1A), ii) highly variable genes and their partners tend to be more highly expressed than lowly variable genes and their partners (Reviewers only Figure 1B) and iii) lowly variable genes and their partners are more likely to be co-expressed than highly variable genes and their partners (Reviewers only Figure 1C). However the effects are not large.

Minor comments:
1. For this sentence "Indeed we find that genetic interaction hubs with variable expression are twice as likely to have gene duplicates as other genetic hubs, and in total have more than six times as many duplicates (Figure 6 and Supplementary Figure 3).", the "in total have more than six times as many duplicates" statement is confusing to me-this seems to imply that the class of variable expression hubs is much larger than the non-variable expression hubs class.

We apologize for the confusion. We removed the phrase “in total have more than six times as many duplicates”.

Reviewer #2:

In this manuscript, Park and Lehner examine the association between genetic interactions and different gene features in order to examine whether genetic interactions could constrain expression variation. The idea is simple and very clever. If Gene A and Gene B show a genetic interaction and their expression is noisy, cells in which gene product A and gene product B are in low abundance will suffer from this genetic interaction, even in wild-type cells. The authors find that in yeast genetic interaction data support their model: genes that have a large number of genetic interactions show less variation in expression levels (noise, plasticity and evolutionary divergence). The model is very interesting (although presented in previous papers by the same author so not completely novel). There are several limitations to the study however.

We are grateful to the reviewer for their positive assessment of our work. Here, we present our responses to the reviewer’s comments and updates to the manuscript, which we hope the reviewer will find satisfactory.

Major comments:
1. It seems that positive and negative genetic interactions should lead to different predictions but these are not considered here. The majority of genetic interactions are negative but the two types should be analyzed separately as to better support the model.

We agree – thank you for this suggestion. Consistent with the model, after correcting for negative genetic interaction degree, there is no significant relationship between positive interaction degree and expression variability (Revised Supplementary Figure 2). In contrast, the correlations between negative genetic interaction degree and gene expression variability remain significant when controlling for positive genetic interaction degree (data not shown). This further supports our model and conclusions.

2. All of the variables that the authors correlate with the number of genetic interactions are also correlated with each other, e.g. noise, plasticity, diversity within species and divergence between species. Thus, it is not clear if all of the analyses support different predictions or are in fact measuring the same thing every time. This is not satisfying and should be dealt with in an appropriate multivariate analysis.

We and others have previously shown that different scales of expression variability (noise, plasticity, evolvability and evolution) are coupled in yeast, presumably because they reflect common underlying molecular mechanisms – see Lehner and Kaneko 2011 for a recent review on this. The

main result that we are presenting here is that genes with more genetic interaction partners have less variable expression and that this is detected at multiple different timescales. Prompted by the Reviewer's comments, we also analyzed partial correlations, e.g. measuring correlation between genetic interaction degree and responsiveness after controlling for the linear effect of noise. The correlations are robust in this analysis, suggesting the possibility of some independence of effects (Revised Supplementary Figure 9). However because the different measures are all subject to technical noise, we would not want to confidently conclude this.

3. The authors seem to imply that there is a causal relationship between the number of genetic interactions and a reduction of expression variation, i.e. genetic interactions constrain the evolution of and regulation of gene expression. The authors do not discuss the possibility that both could be caused by one or several factors that act to constraint or promote these two variables such that there is no causal relationship between the two. For instance, genes with a large number of genetic interactions are involved in a large number of biological processes, which requires that they are stably expressed. Other hypothesis like this could be put forward and test to find the most likely connections.

We have added a sentence to the discussion to acknowledge that the relationship may reflect a co-correlation with a third feature. However, we think that our model is a parsimonious explanation. Multi-functionality (measuring the number of annotated gene ontology terms in Costanzo *et al*., Science 2010) correlates with genetic interaction degree (Revised Supplementary Figure 6A), but only very weakly with different measures of expression variability (Revised Supplementary Figure 6B). Multi-functionality therefore does not account for the relationships between interaction degree and different measures of expression variability (Revised Supplementary Figure 6C).

4. Overall, there is a lack of mechanistic explanation for the trends observed. The yeast data is rich and it should be easily feasible to find examples from the data where a mechanistic basis could be detailed. The manuscript is largely abstract (genetic interactions are indirect measurements of associations between pathways or genes and do not always translate directly into clear mechanisms) and lacks mechanistic explanations.

As reported in the manuscript and in previous work (Lehner, 2010, PLoS Genetics), highly variable expression is mechanistically linked to the lack of a constitutive nucleosome free region in a promoter and the presence of a TATA box. All of these features are avoided in the promoters of genetic interaction hubs, providing a mechanistic explanation for the trends. This data is presented in Figure 5.

5. The correlation coefficients reported in the text are often not the same as those reported in the figures. Correlation coefficients should not be calculated on binned data although binned data can be used for data representation.

We fully agree – correlations are now only reported for unbinned data.

Reviewer #3:

The main finding reported in this paper is that the number of genetic interaction a gene has correlates with the gene expression variability on three time scales: responsiveness to changing conditions, stochastic variations between genetically identical individuals and evolutionary divergence. The authors provide the interesting interpretation to this result, suggesting that reduced expression variability reflects the possible deleterious effect of co-reducing the expression of synthetic-lethal partners, which could limit growth and thereby be selected against. For genes that have multiple synthetic interacting partners, modulating expression might be particularly problematic in this regard. I find this result interesting and worth publication. Yet, I'm not sure it is fully established. Several controls should be considered:

We are grateful to the reviewer for their positive assessment of our work. Here, we present our responses to the reviewer's comments and updates to the manuscript, which we hope the reviewer will find satisfactory.

1. The authors relate only to the number of gene interacting partners. What would be the prediction for interacting pairs? Are they expected to be inversely correlated in expression to reduce the possibility that they will both be low at the same time? Since hubs in the genetic interaction network are characterized by multiple properties (many of which are connected) it will be highly useful to try and support the hypothesis based on distinct (and perhaps less-connected) synthetic interaction pairs.

We thank the reviewer for this suggestion. We analyzed the gene expression level of genetic interacting pairs compared with randomly assigned pairs across different conditions. Consistent with our hypothesis, genetic interacting pairs tend to avoid low gene expression of both partners compared with randomly assigned pairs (Revised Supplementary Figure 8). Thus, the gene expression level of genetically interacting genes could be (somewhat) inversely regulated to reduce detrimental effects.

2. In this context, would the model predict distinct behavior of positive vs. negative interacting partners? This would provide an additional support for their model.

Yes. Please refer to the reply to comment # 1 of reviewer 2 and Revised Supplementary Figure 2.

3. Can the authors suggest how to experimentally test their model? Can they support the notion that (moderate?) down-regulation of synthetic interacting partners reduces growth rate? Do they see evidence for this in the budding yeast literature?

Consistent with the referee's suggestion, genes with more genetic interaction partners are more likely to be haploinsufficient i.e. to cause a growth defect when their dose is reduced by 50% in diploid yeast, (Deutschbauer *et al*., 2005) (Revised Supplementary Figure 7, haploinsufficient genes have twice as many genetic interaction partners as other genes). This enrichment also remains after excluding ribosomal proteins.

How would it fit with the idea that cells are largely robust to quantitative variations in protein levels?

Our argument is that the tight regulation of genetic interaction hubs is one mechanism that confers this robustness to variations in gene expression. This is highlighted in the discussion of the manuscript in the section 'The tight regulation of genetic hubs may contribute to phenotypic robustness'.

4. My understanding is that the analysis is performed on a single dataset only. Although quite large, it is far from being comprehensive and there could be some issues related to the particular genes examined there. There are now many additional datasets available, in particular from the labs of Jonathan Weissman and Kevin Kogan (in addition to the datasets coming from Toronto). The authors should verify that the results extend also to those different datasets.

Following the review's suggestion, we have now added new analyses based on the comprehensive genetic interaction database, BioGRID (Chris Stark, Mike Tyers *et al*., NAR 2011) after removing interactions from the Boone and Andrews labs (i.e. Tong *et al*. and Costanzo *et al*.) (Revised Supplementary Figure 10). The same trends are observed as when analyzing the Costanzo *et al*. dataset highlighting the robustness of our results.

5. It will be nice to discuss the correlates of the synthetic interaction dataset used, as described in the original publication. In particular, the differences between the previously observed correlates (e.g. expression levels, phenotypic capacitance) and the present result should be emphasized. Those previous results might be helpful for the model interpreting the present data as well.

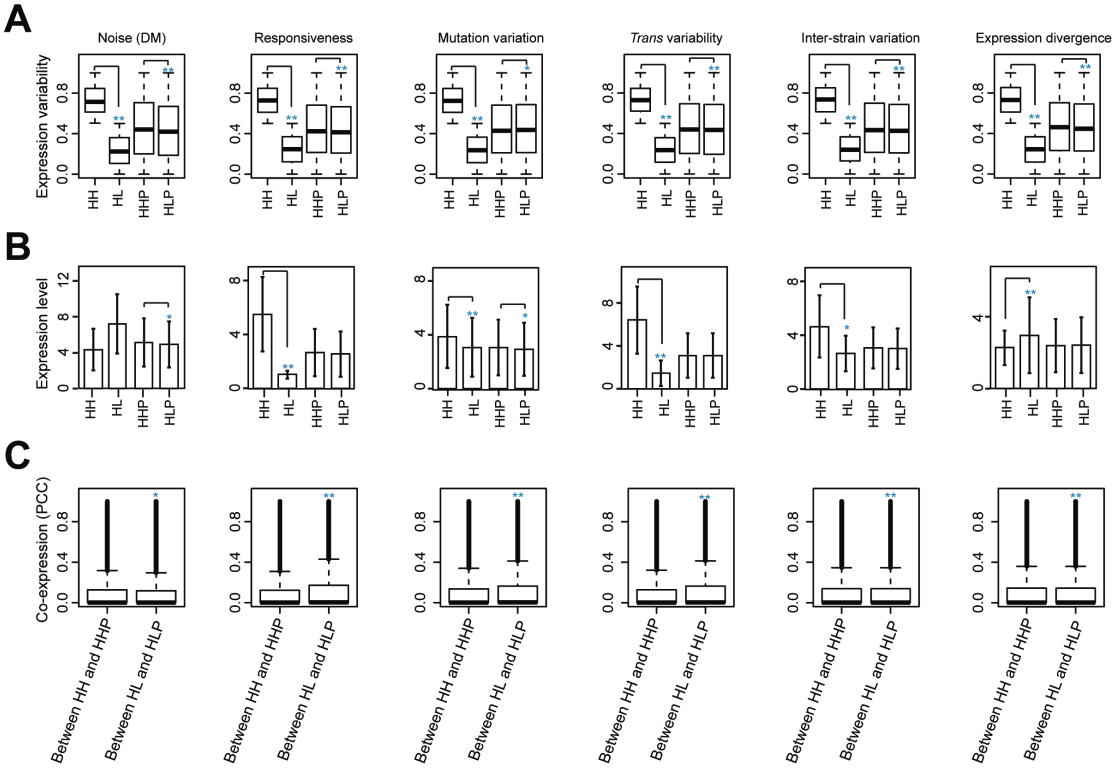
Prompted by the reviewer's comments, we analyzed the correlation between genetic interaction degree and variation in gene expression controlling for the linear effect of gene expression levels or phenotypic capacitance, using partial correlation analysis. While gene expression levels or phenotypic capacitance showed a significant positive correlation as described in the Costanzo *et al*., Science 2010, the correlation between genetic interaction degree and variation in gene expression

remains when controlling for these relationships (Revised Supplementary Figure 3).

6. Correlation strength: In the beginning of the papers, the authors report very low correlation values (<0.2) while at the second half those correlation increase dramatically. Examining the method section, we believe this difference reflects the analysis method: first value corresponds to 'bare' correlation, while the second results from correlating the averaged windows. Now this averaging can artificially increase the correlation values. If this analysis is correct, the authors should be explicit about how they measure correlations and report the real values rather than artificially inflated ones.

We apologize for the confusion. We now only report correlations for the unbinned data and changed the axes labels of the figures.

Reviewers only Figure



Reviewers only Figure 1. Characteristics of the genetic interaction partners of variable and non-variable genes. Each gene is classified into four types: HH; highly variable (variations are above the median value) - highly connected genes (genes with degree at least 25), HL; lowly variable (variations are below the median value) - highly connected genes, HHP; genetic interaction partners of highly variable - highly connected genes and HLP; genetic interaction partners of lowly variable - highly connected genes. (A) Genetic interaction partners of highly (or lowly) variable - highly connected genes are likely to be highly (or lowly) variable. (B) Highly variable – highly connected genes and their partners tend to be more highly expressed than lowly variable – highly connected genes and their partners. (C) Lowly variable – highly connected genes and their partners are more likely to be co-expressed than highly variable – highly connected genes and their partners. *P*-values are calculated using the Mann-Whitney *U* test (**P < 5.0E-3, *P < 5.0E-2).

Acceptance letter 7 January 2013

Thank you again Thank you again for sending us your revised manuscript. We are now satisfied with the modifications made and I am pleased to inform you that your paper has been accepted for publication.

Thank you very much for submitting your work to Molecular Systems Biology.

