molecular
systems
biology

Genomewide landscape of gene–metabolome associations in *Escherichia coli*

Tobias Fuhrer, Mattia Zampieri, Daniel Sévin, Uwe Sauer and Nicola Zamboni

Corresponding author: Uwe Sauer, ETH Zurich

Review timeline:		
	Submission date:	22 June 2016
	Editorial Decision:	25 July 2016
	Revision received:	13 December 2016
	Accepted:	15 December 2016

Editor: Maria Polychronidou

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   25 July 2016

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 study. Overall, the reviewers appreciate the value of the presented resource. However, they raise a number of concerns, which should be carefully addressed in a revision of the manuscript. The reviewers' recommendations are clear so there is no need to repeat all the points listed below. In line with comment #5 of reviewer #1, we would like to ask you to make sure that the metabolomics dataset is well documented and easily accessible.

REFEREE REPORTS

Reviewer #1:

The present manuscript reports what is to my knowledge the 1st genome-wide metabolomics analysis of single gene deletions. The work is conducted in E. coli, a common model organism with a vibrant community of metabolism researchers. The direct infusion metabolomics approach, while subject to many types of interferences, provides the required throughput and is a good match to the genome-scale of the task. Using the resulting data, the authors provide evidence for local effects of gene knockouts on metabolism, not only for enzymes but also for transcription factors (via their enzyme targets) and non-metabolic genes that interact with metabolic genes. They also show strong clustering of related proteins based on their metabolome profiles. And they annotate YhbC is a dihydorotate hydrolase. Overall, the community has been waiting some time for a genome-wide knockout library to be analyzed by metabolomics, and I think there will be great interest in and widespread use of the data.

Major concerns:

1. The metabolomics approach is subject to mis-identification and mis-quantitation, especially due to in-source fragmentation. While this does not detract from the suitability of use of this approach for the present study, the authors need to clearly acknowledge these limations/risks.

2. Many of the figures conveyed essentially nothing to me, either because they were unreadably small or they had no message. In this problem category, I would put 1B, 1C, 1D, 2B, 4A, 4B. The authors need to come up with a compelling set of figures that each convey a clear scientific message (or specific data), as opposed to just a pretty picture.

3. The follow-up work on YhbC is unacceptable quality. The only tangible new knowledge in the paper (separate from the copious and important new data resource) is that YhbC is a dihydorotate hydrolase. Accordingly, this needs to be nailed. The authors minimally need to meet basic biochemistry standards: reporting of specific activities in standard units not random amounts of protein on a random MS scale, showing both depletion of substrate and accumulation of product and whether those are stoichiometric, determining Km, etc. On the intellectual front, this enzyme is quite puzzling, since the cyclization of carbamoyl-aspartate to DHO is thought to be spontaneous (as most intramolecular peptide bond cyclizations are). If this enzyme really catalyzes the reversible reaction, what is the energy source? It is thermodynamically impossible for PyrC and YhbC to catalyze the identical reaction in opposite directions under the same conditions. Can the authors learn anything from when this enzyme is expressed or when it is essential in terms of its physiological role?

4. The final story about deoxycholate was completely lost on me. From the sentence about "mimics gastric stress caused by bile acids" I was confused and I stayed confused through the end of the paper. I cannot say more about the science, because I really do not know what I was supposed to take away from this.

5. I think that the most valuable aspect of this paper should be the resource of metabolomics data for all of the single-gene deletions. In going through the supplement, I did not find this table readily. The quality of this table is paramount to the long-term impact of the paper-- it needs to be easy to find and very well annotated.

Minor issues:

1. I do not love the title, and do not think that the paper is really about "gene-metabolite associations" or "landscapes." It is about metabolomics analysis of a full genome knockout collection.

2. I certainly object to the claims of the paper is unprecedented. What is nice about this paper is that actually it is very well precedented by all sorts of other genome-scale work, especially phenotyping of genome-scale knockout collections, activities that in the past have proven extremely valuable and to make great resources.

3. In Fig 1E, the authors restrict the analysis to genes with at least one significant metabolic change. What fraction of genes showed such a change? If most, then the analysis should be fine, but if it is a minority, the authors probably need to think of a way to avoid such selection bias.

4. While I understand the locality analysis after reading the supplement, I feel that it could have been much better explained in the paper. Looking at figure 2A, I had the idea that all 5 degrees were significant (they aren't radically different from one another), whereas I see in the supplement that the degree 3, 4, 5 are really the control for degree 1, and there is no control to show that degree 5 separation is different or the same as infinite separation.

5. It would be helpful for the authors to provide some examples of non-metabolic enzyme - metabolite assocaitions. What are these typically?

Reviewer #2:

The manuscript by Fuhrer et al. addresses a topical question in systems biology - gene-metabolite (concentration) relationships. To address this, the authors use an unbiased mass-spectrometry approach to characterize metabolomes of the mutant strains in the E. coli single gene knockout library. The results reveal a comprehensive network of gene-metabolite associations that not only recaptures the known and expected links but also presents novel hypotheses, few of which are followed-up in detail. Overall, the study provides a valuable resource and a conceptual step ahead in understanding genotype-(metabolic) phenotype relations. There are, however, a few questions/concerns that need to be addressed.

1. Network proximity analysis: The authors exclude highly connected metabolites from their distance calculations. This approach is not biologically warranted for an unbiased systems study like this (which aims at revealing connections that are mediated through yet unknown mechanisms). Cofactors like NADPH or ATP, for example, can wield a widespread influence on metabolic network connecting distant pathways like PP pathways and TCA cycle. This could potentially explain the discussed cases like malate-aro/pur connection. I therefore recommend performing an additional analysis including physiologically important co-factors in distance calculations. (Use of flux coupling network could also make the analysis more comprehensive).

2. Metabolite annotations: Though authors acknowledge the possible uncertainty in metabolite annotations, the discussion lacks this point. Furthermore, it is unclear whether the pre-processing filters out highly correlated ions that could result from in-source fragmentation or other artifacts.

3. Biological replicates: Perhaps I missed this but what is the R (or rho) value of the responses between the two biological replicates?

4. Semantics:
i) Abstract: "enzyme deletions"
ii) Introduction: "...traits by which genes can influence..." (Traits are not the means by which the genes influence something)

Reviewer #3:

A major challenge, and perhaps the greatest challenge, in the field of microbiology is to better understand the role of the numerous genes of unknown function, even in extensively studied organisms like E. coli. Screening gene knockouts has yielded a great deal of knowledge in this space, but is typically limited by the screening of all gene knockouts against a limited set of factors (e.g., the presence/absence of a specific carbon source). In "Genome-wide landscape of gene-metabolome associations in Escherichia coli" the authors perform metabolomics on thousands of metabolites across the E. coli gene knockout collection to associate genes with metabolic alterations. The result provides a rich collection of information to relate genes to each other and to compare the metabolic changes to environmental stimuli and identify similar metabolic changes associated with gene knockouts. I think the manuscript is a excellent example of how to address this difficult problem of microbial gene annotation. I have only a few minor comments.

1) The utility of this method is somewhat limited by the difficulty of comparing metabolomics data between platforms and laboratories. While in theory a new laboratory would test an additional environmental perturbation and compare their metabolomics result with the dataset from this manuscript, in practice it is quite difficult to do in the context of mass spec, which limits the utility of the generated metabolomics dataset. However, this is simply a limitation of the technology in its current state.

2) It seems the authors use a different culture media in the initial knock-out screen and the later environmental perturbation experiments. I could not find a justification for this switch but it would potentially complicate metabolomics results greatly to have two different background media (e.g., I believe one has casein and the other does not). Although the authors compare the environmental

samples to a control sample with no perturbation on the same media, it is still a little worrisome to switch media like this.

3) prediction of orphan gene
In the main text, this section lacks sufficient methodological information to interpret where they numbers and results come from. While I agree that a lot of the details can be placed in the supplement, I think the balance is a little off here and at least a slightly expanded and intuitive explanation of the methods needs to be brought into the main text.

Typo line 633: annoted should be annotated

1st Revision - authors' response	13 December 2016

Reviewer #1:

The present manuscript reports what is to my knowledge the 1st genome-wide metabolomics analysis of single gene deletions. The work is conducted in E. coli, a common model organism with a vibrant community of metabolism researchers. The direct infusion metabolomics approach, while subject to many types of interferences, provides the required throughput and is a good match to the genome-scale of the task. Using the resulting data, the authors provide evidence for local effects of gene knockouts on metabolism, not only for enzymes but also for transcription factors (via their enzyme targets) and non-metabolic genes that interact with metabolic genes. They also show strong clustering of related proteins based on their metabolome profiles. And they annotate YhbC is a dihydorotate hydrolase. Overall, the community has been waiting some time for a genome-wide knockout library to be analyzed by metabolomics, and I think there will be great interest in and widespread use of the data.

We would like to thank the referee for the positive review and the appreciation of the value of the data for the research community! Please find the point-by-point response below.

Major concerns:

1. The metabolomics approach is subject to mis-identification and mis-quantitation, especially due to in-source fragmentation. While this does not detract from the suitability of use of this approach for the present study, the authors need to clearly acknowledge these limiations/risks.

    In principle we agree with the reviewer that there's a latent risk of mis-quantification and mis-annotation, but not because of spontaneous fragmentation. We have extensive experience and measurements to characterize the extent to fragmentation on pure standards on exactly the same instrument. We observed both in source and post-source fragmentation, i.e. through cooling of ions in the high-pressure collision cell before ions beam focusing and TOF. The hardware is operated and tuned to minimize such problems to only less than a few percent. Hence, degradation fragments are observable only for abundant metabolites. At the level of software, we actually check for common derivatives (loss of P, PPi, CO_2, NH_3, H_2O) throughout the data and use the correlation score between intensities of the molecular ion and that of the putative fragment to check the likelihood that it's indeed a fragment and not a different molecular ion. All of this is convoluted in the annotation score that we report (referenced in the main text as Fuhrer et al 2011). Regardless of this analysis, for any given measured ion we enumerate all possible matches.

    The main issue is mis-quantification and miss-annotation due to peak crowding or presence of unknown unknowns. We added an explicit comment at the very beginning of the results section.

2. Many of the figures conveyed essentially nothing to me, either because they were unreadably small or they had no message. In this problem category, I would put 1B, 1C, 1D, 2B, 4A, 4B. The authors need to come up with a compelling set of figures that each convey a clear scientific message (or specific data), as opposed to just a pretty picture.

    We thank the reviewer for pointing this out. Complex figures were split. We improved the quality of the pictures in terms of size and readability. We also worked over all legends to clarify and explain better what is shown.

3. The follow-up work on YhbC is unacceptable quality. The only tangible new knowledge in the paper (separate from the copious and important new data resource) is that YhbC is a dihydorotate hydrolase. Accordingly, this needs to be nailed. The authors minimally need to meet basic biochemistry standards: reporting of specific activities in standard units not random amounts of protein on a random MS scale, showing both depletion of substrate and accumulation of product and whether those are stoichiometric, determining Km, etc. On the intellectual front, this enzyme is quite puzzling, since the cyclization of carbamoyl-aspartate to DHO is thought to be spontaneous (as most intramolecular peptide bond cyclizations are). If this enzyme really catalyzes the reversible reaction, what is the energy source? It is thermodynamically impossible for PyrC and YhbC to catalyze the identical reaction in opposite directions under the same conditions. Can the authors learn anything from when this enzyme is expressed or when it is essential in terms of its physiological role?

    We appreciate the comment that highlighted a superficial discussion of the results. The apparent confusion originates from the fact that the in vitro the reaction can be driven in either direction (without any cofactor) by excess substrate. In vivo, the pathway is likely to work in the canonical direction.

    Overall, we regret that we can't provide the information to the biochemical detail outlines by the Reviewer (Km etc). One single, very detailed example is of little relevance and doesn't change much in the economy of a genome-wide study. As highlighted in the text and by the reviewers, the main contribution is the resource and the analysis that points to massive presence of "distal" gene-metabolite interactions. Unfortunately, the molecular origin of such links remains largely unexplained at the moment, but the connections are statistically significant.

    The function predictions and stress response are examples on how the association map can be queried to generate hypothesis on gene function. Because of the nature of the data, the complexity of the response, and the incomplete metabolome coverage, the prediction related to the pathway or functional level. This aspect is reflected in the detailed discussion of the 70+ prediction that we inferred using the CLR algorithm. To avoid a misrepresentation of the results and merits of the prediction, we decided to remove the incomplete analysis of the YhbC case.

4. The final story about deoxycholate was completely lost on me. From the sentence about "mimics gastric stress caused by bile acids" I was confused and I stayed confused through the end of the paper. I cannot say more about the science, because I really do not know what I was supposed to take away from this.

    We thank the reviewer for pointing out that this part was unclear. We have now clarified the paragraph about the metabolic response to environmental perturbations and in particular to deoxycholate stress.

5. I think that the most valuable aspect of this paper should be the resource of metabolomics data for all of the single-gene deletions. In going through the supplement, I did not find this table readily. The quality of this table is paramount to the long-term impact of the paper-- it needs to be easy to find and very well annotated.

    We thank the reviewer and editor for pointing this out and we totally agree that accessibility of the metabolomics data is crucial. The information of how and where to

access the raw data online is now contained in the main text, at the end of Materials and Methods in the new section "Data availability".

Minor issues:

1. I do not love the title, and do not think that the paper is really about "gene-metabolite associations" or "landscapes." It is about metabolomics analysis of a full genome knockout collection.

   Apparently the reviewer prefers a title that describes what was done, whereas we prefer a title that describes what was achieved. The main achievement in our view is the extensive map of gene-metabolite associations that will be made available to the community. The summary of these interactions is actually a map of a landscape that we never looked at in this unbiased fashion. Hence, we prefer to stick to our title. We do describe very clearly what was done in the abstract.

2. I certainly object to the claims of the paper is unprecedented. What is nice about this paper is that actually it is very well precedented by all sorts of other genome-scale work, especially phenotyping of genome-scale knockout collections, activities that in the past have proven extremely valuable and to make great resources.

   We never intended to claim that we performed the first genome-scale analysis of any type, but we are not aware of another metabolomics study at this scale. Unprecedented refered to the fact that phenotypic studies generally have one or very few read outs (such as growth rate), while we provide multiple readouts for each case. In some ways the reviewer even supports our claim by saying that the field has ben waiting for this. Nevertheless, we removed the term entirely and revised the text for clarification.
   The key advantage of our metabolomics approach over previous phenotypic studies is the capacity to detect responses even in the absence of growth phenotypes, thereby increasing functional readouts by orders of magnitude compared to classical phenotypic screens. In this sense, our compendium of >7000 relative metabolite ion levels in >3800 mutants is unprecedented.

3. In Fig 1E, the authors restrict the analysis to genes with at least one significant metabolic change. What fraction of genes showed such a change? If most, then the analysis should be fine, but if it is a minority, the authors probably need to think of a way to avoid such selection bias.

   Good point. If we apply a significance cutoff of absolute(z-score) > 5, we find that about 50% of the gene knock-outs exhibited at least one metabolic change. Hence, the analysis was based on considering half of the genes and is therefore representative.

4. While I understand the locality analysis after reading the supplement, I feel that it could have been much better explained in the paper. Looking at figure 2A, I had the idea that all 5 degrees were significant (they aren't radically different from one another), whereas I see in the supplement that the degree 3, 4, 5 are really the control for degree 1, and there is no control to show that degree 5 separation is different or the same as infinite separation.

   Notably, the larger the distance between enzyme and metabolite, the fewer pairs can be retrieved. We here show the statistics up to a distance of five because we need a sufficient number of enzyme-metabolite pairs to draw a reliable estimate of p-value distribution. Moreover, the distribution of p-values already shows clearly that after a distance of 2 there is no enrichment of larger metabolic changes in the proximity of the deleted enzyme, and distribution of p values become uniform.

5. It would be helpful for the authors to provide some examples of non-metabolic enzyme - metabolite assocaitions. What are these typically?

   We apologize for the confusion. What we meant are non-enzymatic gene-metabolite associations based on non-metabolic genes listed in the ECID database (see Figure EV6C).

Examples of typical non-metabolic associations include maturation factors (e.g. YgfY, recently annotated as SdhE) affecting the reactant levels of the respective enzyme function, transporters (e.g. brnQ, branched chain amino acid transporter) affecting intermediates of the respective degradation pathways or transcriptional regulators (e.g. puuR, DNA binding transcriptional repressor sensing putrescine). The examples were added to Figure 1 in the case of BrnQ and PuuR while YgfY is reported in Figure EV8.

Reviewer #2:

The manuscript by Fuhrer et al. addresses a topical question in systems biology - gene-metabolite (concentration) relationships. To address this, the authors use an unbiased mass-spectrometry approach to characterize metabolomes of the mutant strains in the E. coli single gene knockout library. The results reveal a comprehensive network of gene-metabolite associations that not only recaptures the known and expected links but also presents novel hypotheses, few of which are followed-up in detail. Overall, the study provides a valuable resource and a conceptual step ahead in understanding genotype-(metabolic) phenotype relations. There are, however, a few questions/concerns that need to be addressed.

Firstly, we would like to thank the referee for the positive feedbacks and comments. Please find the point-by-point response below.

1. Network proximity analysis: The authors exclude highly connected metabolites from their distance calculations. This approach is not biologically warranted for an unbiased systems study like this (which aims at revealing connections that are mediated through yet unknown mechanisms). Cofactors like NADPH or ATP, for example, can wield a widespread influence on metabolic network connecting distant pathways like PP pathways and TCA cycle. This could potentially explain the discussed cases like malate-aro/pur connection. I therefore recommend performing an additional analysis including physiologically important co-factors in distance calculations. (Use of flux coupling network could also make the analysis more comprehensive).

   Exclusion of cofactors or 'currency metabolites' is quite common in the field and was shown to be necessary to obtain meaningful predictions (Hancock et al, 2012; Kim et al, 2015; Noor et al, 2010; Ravasz et al, 2002; Schulz et al, 2014). This is compatible with the view that removing a reaction that uses ATP, NADPH, or the C1 pool is unlikely to affect other reactions that use the same cofactor since it's availability isn't compromised. To our knowledge, cross-pathway interactions mediated by cofactors are mostly allosteric and not catalytic (e.g. NADPH-repression of the Zwf, ATP binding to Pfk).

   This is different if the deletion leads to a reduction of the "active" form of the cofactor. This happens if the main cellular processes responsible for cofactor production or regeneration are disturbed. In our data, we observed that gene deletions in coenzyme transport and metabolism, nucleotide transport and metabolism, signal transduction mechanisms and transcription tend to induce numerous metabolic changes (new Figure EV7). This indicates that all these processes are important and mutations lead to pleiotropic effects. This isn't surprising, because such mutations prevent the generation of the cofactors, which likely becomes limiting for all coupled reactions. The metabolic consequences of cofactor depletion are widespread. They don't explain the emergence of rather specific association between KOs and apparently distal metabolites.

   We now included a new figure EV7, and discussed this aspect in the main text.

   References:

   Hancock T, Wicker N, Takigawa I, Mamitsuka H (2012) Identifying neighborhoods of coordinated gene expression and metabolite profiles. PloS one 7: e31345

Kim T, Dreher K, Nilo-Poyanco R, Lee I, Fiehn O, Lange BM, Nikolau BJ, Sumner L, Welti R, Wurtele ES, Rhee SY (2015) Patterns of metabolite changes identified from large-scale gene perturbations in Arabidopsis using a genome-scale metabolic network. Plant physiology 167: 1685-1698

Noor E, Eden E, Milo R, Alon U (2010) Central carbon metabolism as a minimal biochemical walk between precursors for biomass and energy. Mol Cell 39: 809-820

Ravasz E, Somera AL, Mongru DA, Oltvai ZN, Barabasi AL (2002) Hierarchical organization of modularity in metabolic networks. Science 297: 1551-1555

Schulz JC, Zampieri M, Wanka S, von Mering C, Sauer U (2014) Large-scale functional analysis of the roles of phosphorylation in yeast metabolic pathways. Sci Signal 7: rs6

2. Metabolite annotations: Though authors acknowledge the possible uncertainty in metabolite annotations, the discussion lacks this point. Furthermore, it is unclear whether the pre-processing filters out highly correlated ions that could result from in-source fragmentation or other artifacts.

    We thank the reviewer for pointing this out. We added a sentence discussing the impact of potential miss-annotations. In the published method paper (referenced in the main text as Fuhrer et al 2011) we investigated generally observed highly correlating ions and identified the respective frequent mass shifts mostly as neutral losses such as for example H_2O or NH_3 as well as adducts such as $+H_2PO_4Na$ or $+H_2PO_4K$. These frequently occurring mass shifts are included in the annotation procedure.

    In addition the voltage settings of the mass-spectrometry method were chosen to prevent in-source fragmentation and other artifacts as much as possible.

3. Biological replicates: Perhaps I missed this but what is the R (or rho) value of the responses between the two biological replicates?

    The mean rho is 0.1059. This information wasn't included because rho is not a good benchmark for reproducibility given that we measured thousands of features but only a little fraction is significantly associated to a given genotype. The effect of metabolic "markers" is eclipsed by the noise associated to the overwhelming number of non-markers and z-score normalization. To bypass this issue, we assessed reproducibility by estimating a background distribution of Z-scores using the biological replicates. The analysis and results are now reported in figure EV2A.

4.   Semantics:

    i) Abstract: "enzyme deletions"

    Corrected.

    ii) Introduction: "...traits by which genes can influence..." (Traits are not the means by which the genes influence something)

    Corrected.

Reviewer #3:

A major challenge, and perhaps the greatest challenge, in the field of microbiology is to better understand the role of the numerous genes of unknown function, even in extensively studied organisms like E. coli. Screening gene knockouts has yielded a great deal of knowledge in this space, but is typically limited by the screening of all gene knockouts against a limited set of factors

(e.g., the presence/absence of a specific carbon source). In "Genome-wide landscape of gene-metabolome associations in Escherichia coli" the authors perform metabolomics on thousands of metabolites across the E. coli gene knockout collection to associate genes with metabolic alterations. The result provides a rich collection of information to relate genes to each other and to compare the metabolic changes to environmental stimuli and identify similar metabolic changes associated with gene knockouts. I think the manuscript is a excellent example of how to address this difficult problem of microbial gene annotation. I have only a few minor comments.

We thank the reviewer for the very positive review. Please see below for the point-by-point responses.

1. The utility of this method is somewhat limited by the difficulty of comparing metabolomics data between platforms and laboratories. While in theory a new laboratory would test an additional environmental perturbation and compare their metabolomics result with the dataset from this manuscript, in practice it is quite difficult to do in the context of mass spec, which limits the utility of the generated metabolomics dataset. However, this is simply a limitation of the technology in its current state.

   The preconditions aren't overly strict, because (in our experience) it's sufficient to compare qualitative changes for features that can be matched with deprotonated/protonated metabolites. We already have done (successfully) this with data generated on different MS instruments that offer a similar or better coverage of metabolism and with cells grown in different media.

2. It seems the authors use a different culture media in the initial knock-out screen and the later environmental perturbation experiments. I could not find a justification for this switch but it would potentially complicate metabolomics results greatly to have two different background media (e.g., I believe one has casein and the other does not). Although the authors compare the environmental samples to a control sample with no perturbation on the same media, it is still a little worrisome to switch media like this.

   We thank the reviewer for raising the point on potential drawbacks from comparing different. In the screening, we opted for a rich medium with casein amino acids to (i) better reflect the variety of natural substrates, (ii) compensate auxotrophies of mutants lacking genes that are essential for growth on glucose, (iii) alleviate growth rate differences. The exact composition of the screening medium isn't known. For instance, it also contains nucleobases (xanthine and hypoxanthine) and other peaks we can't assign.

   The follow ups were done with glucose minimal medium to reduce potential confounding variables. This is common practice, but could be a source of difference in results. For the cases we discussed, this change doesn't seem to be relevant. It is also plausible that in our examples of environmental perturbations such as phosphate or sulphur limitation, the presence of amino acids would not bias our predictions. These results actually speak in favor of the resource we generated. We expect that for certain questions involving amino acids metabolism one would have to use a rich medium to compare responses. We now clarify this limitation in the text.

3. prediction of orphan gene: In the main text, this section lacks sufficient methodological information to interpret where they numbers and results come from. While I agree that a lot of the details can be placed in the supplement, I think the balance is a little off here and at least a slightly expanded and intuitive explanation of the methods needs to be brought into the main text.

   Also requested by the Editor, all Materials and Methods as well as Results from the Supplements have been moved to the main text and merged wherever necessary.

4. Typo line 633: annoted should be annotated

Corrected

2nd Editorial Decision   15 December 2016

Thank you for sending us your revised manuscript. We have now heard back from reviewer #2 who was asked to evaluate the study. The reviewer is satisfied with the modifications made and I am pleased to inform you that your paper has been accepted for publication.

Reviewer #2:

The authors have addressed my comments in the revised version.

Corrected

Corresponding Author Name: Uwe Sauer	
Journal Submitted to: MSB	
Manuscript Number: MSB-16-7150R	

Reporting Checklist For Life Sciences Articles (Rev. July 2015)

This checklist is used to ensure good reporting standards and to improve the reproducibility of published results. These guidelines are consistent with the Principles and Guidelines for Reporting Preclinical Research issued by the NIH in 2014. Please follow the journal's authorship guidelines in preparing your manuscript.

A- Figures

1. Data

The data shown in figures should satisfy the following conditions:

➔ the data were obtained and processed according to the field's best practice and are presented to reflect the results of the experiments in an accurate and unbiased manner.
➔ figure panels include only data points, measurements or observations that can be compared to each other in a scientifically meaningful way.
➔ graphs include clearly labeled error bars for independent experiments and sample sizes. Unless justified, error bars should not be shown for technical replicates.
➔ if n< 5, the individual data points from each experiment should be plotted and any statistical test employed should be justified
➔ Source Data should be included to report the data underlying graphs. Please follow the guidelines set out in the author ship guidelines on Data Presentation.

2. Captions

Each figure caption should contain the following information, for each panel where they are relevant:

➔ a specification of the experimental system investigated (eg cell line, species name).
➔ the assay(s) and method(s) used to carry out the reported observations and measurements
➔ an explicit mention of the biological and chemical entity(ies) that are being measured.
➔ an explicit mention of the biological and chemical entity(ies) that are altered/varied/perturbed in a controlled manner.

➔ the exact sample size (n) for each experimental group/condition, given as a number, not a range;
➔ a description of the sample collection allowing the reader to understand whether the samples represent technical or biological replicates (including how many animals, litters, cultures, etc.).
➔ a statement of how many times the experiment shown was independently replicated in the laboratory.
➔ definitions of statistical methods and measures:
  • common tests, such as t-test (please specify whether paired vs. unpaired), simple χ2 tests, Wilcoxon and Mann-Whitney tests, can be unambiguously identified by name only, but more complex techniques should be described in the methods section;
  • are tests one-sided or two-sided?
  • are there adjustments for multiple comparisons?
  • exact statistical test results, e.g., P values = x but not P values < x;
  • definition of 'center values' as median or average;
  • definition of error bars as s.d. or s.e.m.

Any descriptions too long for the figure legend should be included in the methods section and/or with the source data.

Please ensure that the answers to the following questions are reported in the manuscript itself. We encourage you to include a specific subsection in the methods section for statistics, reagents, animal models and human subjects.

B- Statistics and general methods

Please fill out these boxes ↓ (Do not worry if you cannot see all your text once you press return)

Question	Answer
1.a. How was the sample size chosen to ensure adequate power to detect a pre-specified effect size?	Two biological replicates (independent clones) were extracted and measured with two technical replicates.
1.b. For animal studies, include a statement about sample size estimate even if no statistical methods were used.	not applicable
2. Describe inclusion/exclusion criteria if samples or animals were excluded from the analysis. Were the criteria pre-established?	Samples were excluded from analysis based on extremely sick growth phenotypes and non-reproducible effects (based on z-score). See Physiology paragraph and Data normalization and calculation of differential ions paragraph in Material and Methods.
3. Were any steps taken to minimize the effects of subjective bias when allocating animals/samples to treatment (e.g. randomization procedure)? If yes, please describe.	Sample acquistion was randomized.
For animal studies, include a statement about randomization even if no randomization was used.	not applicable
4.a. Were any steps taken to minimize the effects of subjective bias during group allocation or/and when assessing results (e.g. blinding of the investigator)? If yes please describe.	not applicable
4.b. For animal studies, include a statement about blinding even if no blinding was done	not applicable
5. For every figure, are statistical tests justified as appropriate?	Yes.
Do the data meet the assumptions of the tests (e.g., normal distribution)? Describe any methods used to assess it.	Yes.
Is there an estimate of variation within each group of data?	not applicable
Is the variance similar between the groups that are being statistically compared?	Yes, see data normalization and calculation of differential ions paragraph in Material and Methods.

C- Reagents

6. To show that antibodies were profiled for use in the system under study (assay and species), provide a citation, catalog number and/or clone number, supplementary information or reference to an antibody validation profile. e.g., Antibodypedia (see link list at top right), 1DegreeBio (see link list at top right).	not applicable
7. Identify the source of cell lines and report if they were recently authenticated (e.g., by STR profiling) and tested for mycoplasma contamination.	not applicable

* for all hyperlinks, please see the table at the top right of the document

D- Animal Models

8. Report species, strain, gender, age of animals and genetic modification status where applicable. Please detail housing and husbandry conditions and the source of animals.	not applicable
9. For experiments involving live vertebrates, include a statement of compliance with ethical regulations and identify the committee(s) approving the experiments.	not applicable
10. We recommend consulting the ARRIVE guidelines (see link list at top right) (PLoS Biol. 8(6), e1000412, 2010) to ensure that other relevant aspects of animal studies are adequately reported. See author guidelines, under 'Reporting Guidelines'. See also: NIH (see link list at top right) and MRC (see link list at top right) recommendations. Please confirm compliance.	not applicable

E- Human Subjects

11. Identify the committee(s) approving the study protocol.	not applicable
12. Include a statement confirming that informed consent was obtained from all subjects and that the experiments conformed to the principles set out in the WMA Declaration of Helsinki and the Department of Health and Human Services Belmont Report.	not applicable
13. For publication of patient photos, include a statement confirming that consent to publish was obtained.	not applicable
14. Report any restrictions on the availability (and/or on the use) of human data or samples.	not applicable
15. Report the clinical trial registration number (at ClinicalTrials.gov or equivalent), where applicable.	not applicable
16. For phase II and III randomized controlled trials, please refer to the CONSORT flow diagram (see link list at top right) and submit the CONSORT checklist (see link list at top right) with your submission. See author guidelines, under 'Reporting Guidelines'. Please confirm you have submitted this list.	not applicable
17. For tumor marker prognostic studies, we recommend that you follow the REMARK reporting guidelines (see link list at top right). See author guidelines, under 'Reporting Guidelines'. Please confirm you have followed these guidelines.	not applicable

F- Data Accessibility

18. Provide accession codes for deposited data. See author guidelines, under 'Data Deposition'. Data deposition in a public repository is mandatory for: a. Protein, DNA and RNA sequences b. Macromolecular structures c. Crystallographic data for small molecules d. Functional genomics data e. Proteomics and molecular interactions	The following data are available as separate files online: Profile data for > 34'000 mass spectrometric analysis can be downloaded from http://massive.ucsd.edu/, Accession MSV000078963.
19. Deposition is strongly recommended for any datasets that are central and integral to the study; please consider the journal's data policy. If no structured public repository exists for a given data type, we encourage the provision of datasets in the manuscript as a Supplementary Document (see author guidelines under 'Expanded View' or in unstructured repositories such as Dryad (see link list at top right) or Figshare (see link list at top right).	Raw data and modified z-scores for positive and negative mode (tab-separated and excel files) can be downloaded from http://www.imsb.ethz.ch/research/sauer/resources.html.
20. Access to human clinical and genomic datasets should be provided with as few restrictions as possible while respecting ethical obligations to the patients and relevant medical and legal issues. If practically possible and compatible with the individual consent agreement used in the study, such data should be deposited in one of the major public access-controlled repositories such as dbGAP (see link list at top right) or EGA (see link list at top right).	not applicable
21. As far as possible, primary and referenced data should be formally cited in a Data Availability section. Please state whether you have included this section. Examples: **Primary Data** Wetmore KM, Deutschbauer AM, Price MN, Arkin AP (2012). Comparison of gene expression and mutant fitness in Shewanella oneidensis MR-1. Gene Expression Omnibus GSE39462 **Referenced Data** Huang J, Brown AF, Lei M (2012). Crystal structure of the TRBD domain of TERT and the CR4/5 of TR. Protein Data Bank 4O26 AP-MS analysis of human histone deacetylase interactions in CEM-T cells (2013). PRIDE PXD000208	not applicable
22. Computational models that are central and integral to a study should be shared without restrictions and provided in a machine-readable form. The relevant accession numbers or links should be provided. When possible, standardized format (SBML, CellML) should be used instead of scripts (e.g. MATLAB). Authors are strongly encouraged to follow the MIRIAM guidelines (see link list at top right) and deposit their model in a public database such as Biomodels (see link list at top right) or JWS Online (see link list at top right). If computer source code is provided with the paper, it should be deposited in a public repository or included in supplementary information.	not applicable

G- Dual use research of concern

23. Could your study fall under dual use research restrictions? Please check biosecurity documents (see link list at top right) and list of select agents and toxins (APHIS/CDC) (see link list at top right). According to our biosecurity guidelines, provide a statement only if it could.	not applicable