molecular systems biology

Reverse-engineering a hierarchical regulatory network downstream of oncogenic KRAS

Iwona Stelniec-Klotz, Stefan Legewie, Oleg Tchernitsa, Franziska Witzel, Bertram Klinger, Christine Sers, Hanspeter Herzel, Nils Blüthgen and Reinhold Schäfer

Corresponding author: Reinhold Schäfer, Charité University Medicine Berlin

Review timeline:

Submission date:	06 December 2010
Editorial Decision:	01 February 2010
Re-submission:	23 April 2012
Editorial Decision:	23 May 2012
Revision received:	26 June 2012
Accepted:	29 June 2012

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 01 February 2010

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 raise substantial concerns on your work, which, I am afraid to say, preclude its publication.

The reviewers recognized that a better understanding of KRAS-dependent gene regulation in cancer would be potentially valuable, but they indicated clearly that the experiments conducted in this work were currently insufficient to rigorously validate the predicted KRAS-dependent regulatory network. Most importantly, the second reviewer noted that some of the model's predicted regulatory interactions run contrary to existing literature reports, and therefore they deserve direct experimental testing. Moreover, all three reviewers found the existing experimental validation unconvincing, with reviewer #1 raising concerns that the small set of genes used in the transcriptional profiling experiments could potentially bias the results, and the second reviewer indicating that more comprehensive phenotypic analyses would be needed to fully validate the model predictions. Overall, these concerns raised important doubts about the conclusiveness and physiological relevance of the regulatory logic in this proposed KRAS network.

Given these clear concerns, we feel we must return this work to you with the message that we cannot offer to publish it.

Nevertheless, the reviewers did express interest in this work, and they indicated that extensive additional experiments may be able to provide the needed level of support and validation. As such, we may be willing to reconsider a new submission based on this work. Such a work would need to

© European Molecular Biology Organization

include new experiments that directly support and validate the proposed regulatory network. The reviewers recommend more extensive microarray or phenotypic analyses, as well as direct testing of some of the novel (and potentially surprising) regulatory interactions. In addition, the editor would like to suggest that if this model could be used to predict the outcomes of combinatorial perturbations or knockdowns, that would certainly help to support the practical value of this network reconstruction approach. We appreciate that this would require substantial additional experimental work, and we would understand if you decided, instead, to submit this work to another journal. A new submission would have a new number and receipt date. If you do decide to follow this course then please enclose with your re-submission an account of how the work has been altered in response to the points raised in the present review.

I am sorry that the review of your work did not result in a more favorable outcome on this occasion, but I hope that you will not be discouraged from sending your work Molecular Systems Biology in the future.

Also, I acknowledge that the review process took somewhat longer than usual in this case, due to delays in receiving the reviewers' reports.

Thank you for the opportunity to examine this work.

Yours sincerely,

Editor
Molecular Systems Biology

Reviewer #1 (Remarks to the Author):

The paper by Stelniec et al aims to elucidate regulatory networks downstream of RAS in ovarian cancer in a combined experimental-computational approach.

The study targets 7 transcription factors (TFs) by RNAi and reconstructs a regulatory network from perturbation effects. The main data set consists of changes in the 7 TFs observed after perturbing one of them. From this data the authors reconstruct a regulatory network showing a two-layered regulatory hierarchy.

Additional evidence for this hierarchy comes from the analysis of growth phenotypes showing the responses to perturbing members of the upper layer are a superset of the responses to perturbing members of the lower layer. An interpretation would be that the lower layer has a more specialized regulatory than the upper layer.

A third data set analyzed shows the transcriptional response of 329 validated RAS pathway target genes to TF perturbations. This part of the analysis is somewhat weaker than the rest. What is the scientific rationale for restricting the analysis to this small set of genes? It's good to have a list of validated targets, but a genome-wide analysis could have shown much more overlap between transcriptional profiles. The observations that patterns of target regulation did not overlap substantially could be completely due to this bias in gene selection. In particular, if the overlap is so small, doesn't that contradict the tight connections we see within the layers in Fig 4? Why are the two layers not already visible in Fig 2D, which shows the overlap of transcriptional responses?

The rationale behind the growth data is "responses mediated by downstream factors ... were a subset of those mediated by upstream factors". Shouldn't this subset relation also be visible in the expression data? Genes regulated by the downstream factors should also react if the upstream factors are perturbed, while genes directly regulated by the upstream factors should only respond to perturbing the upstream factors.

I don't expect all of this to be visible 100%. I understand about noise, feedback and compensation mechanisms. But the low overlap between transcriptional responses worries me, given that the main result of the paper is a highly connected network.

© European Molecular Biology Organization

Reviewer #2 (Remarks to the Author):

In this manuscript, the authors adopt a systems biology approach to dissect the regulatory network involving 7 transcription factors, selected from 51 TFs up-regulated by the KRAS oncoprotein in an in vitro transformation system of rat ovarian epithelial cells. On the basis of a reverse-engineering approach applied to the results generated by perturbation experiments (selective inhibition of signaling pathways and knockdown of each of the 7 TFs followed by RT-PCR and WB analysis), the authors propose a network in which the analyzed transcription factors can be separated in two hierarchically distint groups. In support of the proposed model, the authors show the effect of the members of the two groups of TFs on several growth parameters and target gene expression.

Given the relevance of the KRAS-dependent transcriptional networks in human tumorigenesis, this is a very interesting contribution, based on a strategy (MRA, modular response analysis) previously applied to the study of various gene networks. However, the experimental validation of the model prediction is incomplete and not fully convincing.

Major points:

1. The inferred regulatory network (Fig.3B and scheme in Fig.4) leads to predictions concerning:
a) the reciprocal interactions between between 3 signaling pathways and 7 transcription factors (panels 1,4,7,9 in Fig.3B);
b) the reciprocal interactions between the 7 transcription factors (panels 3 in Fig.3B), leading to the distinction between two hierarchical groups (Fig-4).
While the proposed transcription factor hierarchy is tested in Fig.5, the model predictions of the effect of signaling pathways on protein levels of TFs are not experimentally verified. This would be particularly relevant in the case of Fosl1, for which the hitherto unknown model prediction is in striking contrast with multiple independent reports (Vial E. et al., J. Cell Sci. 2003; Casalino et al., MCB 2003; Murphy LO et al., MCB 2004;Basbous et al., 2007) showing that ERK phosphorylation positive controls the level of the Fosl1 gene product (Fra-1).
To test the reliability of the proposed model, the authors should show that in their cell system Fra-1 is actually negatively controlled by ERK activity (by investigating, for example, the Fra-1 stability and/or accumulation of the protein ectopically expressed under control of a heterologous promoter, as shown in Basbous et al., 2007).

2. Along with the gene expression clusters inferred by expression profiling of the transcripts affected by each of the 7 transcription factors, the authors suggest that the differential effect on 2D growth of RelA/Otx1/Gfi1 vs Hmga2/JunB/Fosl1/Klf6 (Fig.5) reflects the proposed hierarchy between "upstream" and "downstream" transcription factors (Fig.4).
For full validation of the model prediction, the authors should test if other parameters of the KRAS-dependent neoplastic transformation, such as cell motility, invasiveness, anoikis, etc., are differentially affected by the knockdown of members of the two "hierarchically separated" groups of transcription factors.

Reviewer #3 (Remarks to the Author):

The authors apply a systems approach to investigate ROSE-RAS cells. They measure mRNA expression with a customized array, knock-down specific nodes, transcription factors and cell signaling components, and measure outcome phenotype with invasion assays. There experimental procedures appear to be valid but the computational analysis does not appear fully cohesive and to the point. There are many small but important issues that need to be address before the paper can be published.

Specific comments:

"So far, signaling kinase inhibitors which target the PI-3-K/Akt- or Raf/Mek/Erk-pathways have shown limited therapeutic efficacy in the clinical setting." Please elaborate with references.

"Microarray analysis permits the simultaneous assessment of many network elements. However, the derivation of the network topology from microarray data is not straight-forward, as network perturbation data are lacking". Most microarray analyses are done on perturbation data. The problem

© European Molecular Biology Organization

with microarrays is that they measure mRNA levels which do not directly reflect regulatory mechanisms. This is not clearly stated by the authors.

"Bayesian statistics, neural network type ODE models (Busch et al, 2009) and metabolic control analysis (Kholodenko et al, 2002)" More examples are needed to show appropriate coverage of the topic. Tools such as ARACNE the infderelator and eureqa should be mentioned.

Page 4 "contrasted" I think it was meant "constructed"

"the adhesion molecule E-cadherin (Malek et al., unpublished data)." If microarrays were performed and the data was previously published why there is a reference to Malek et al., unpublished data?

Explain "anoikis" for the non-expert reader.

Page 4 "Reversely" should be "reverse"

Page 4 "focussed" should be "focused"

"we focussed on the transcription factors Fosl1, Gfi1, Hmga2, JunB, Klf6, and Otx1" Is it known that these factors regulate the genes that changed in expression? Do the factors have binding motif sequences at the promoters of the genes that changed in expression? Is there ChIP evidence for any of these factors? Can the kinase be linked to the factors through known pathways?

"RelA was not included in the KRAS pathway-dependent gene signature, but was up-regulated at the protein level" protein level experiments never introduced before this comment. The reader does know about these experiments yet.

"modular response analysis" approach should be explained at the bottom of page 4 and not so much in the results.

Most of page 6 belongs in the introduction.

Direct physical interaction between RelA and Hmga2 should be mentioned (PMID: 12693954)

"customised microarrays representing 329 independently validated RAS pathway-target genes. KRAS oncogene-mediated transformation of ROSE cells was associated with the deregulation of 147 target genes." Not clear if the 147 are a subset of the 329? If the 147 are a subset of the 329, why the rest of the genes are on the microarray? Need clear and detailed explanation.

"We concluded that the RAS pathway orchestrated the de-regulation of the transcriptome via a limited number of transcription factors, each of which controls the transcription of distinct subsets of target genes." This conclusion is problematic as there could be many other explanations for the observed results. Randomized controls need to be created. Also, there is no mention of genes that were aggravated by the knockdown, i.e. further increased or further decrease in expression due to the knock-down.

Figure 3A is most central to the paper but difficult to understand. From the figure it looks like inhibiting pErk does not have an effect or has an effect that is similar to the over-expression of Ras. Is this true? It is also difficult to understand how the MRA algorithm came to the conclusions it did given the results from 3A. The colors do not match.

"calculated the correlation coefficients between transcription factor and target gene mRNAs over 9 experimental perturbation conditions" Not clear why this was done this way and not just compare the expression of the genes fold change or just change vs. the control. There is not mention in the methods how the correlation coefficients were calculated.

The way the global response matrix R depicted in Figure 3A is too complex and not producing clear results. It should be replaced with a simpler procedure.

"feedback regulation of exogenously overexpressed RAS protein by other nodes in the network" Not

© European Molecular Biology Organization

clear. Why not shown?

"The best model fit agrees well with the measured perturbationresponse data (Fig. 3C)." This needs to be further quantified and explained.

The reduced growth by knocking down the transcription factors is not a valid validation of the model by itself since such experiments could have been done regardless of the modeling.

The MRA model is relatively complex so one wonders if the separation between Otx1, Gfi1 and RelA vs. Hmga2, Fosl1, Klf6 and JunB that was shown for the validation in Fig. 5 could have been identified by correlation analysis.

What is the relationship between the transcription factors within each layer?

In figure 5E legend and text it says 7 + 2 = 9 conditions but only 7 are shown, what about the Akt and Erk conditions? How come these are not a part of the clustergram?

Re-submission 23 April 2012

Thank you very much for expressing an interest in our work. We very much appreciate that you are willing to consider an advanced version of our manuscript as a new submission. As suggested by the reviewers, we have included additional microarray experiments, functional and phenotypic assays. In addition, we followed your suggestion to test the model with double-perturbation experiments. Please find enclosed an account of how we have completed our work in response to the points raised in the previous review.

Remarks of Reviewer #1: *The paper by Stelniec et al aims to elucidate regulatory networks downstream of RAS in ovarian cancer in a combined experimental-computational approach. The study targets 7 transcription factors (TFs) by RNAi and reconstructs a regulatory network from perturbation effects. The main data set consists of changes in the 7 TFs observed after perturbing one of them. From this data the authors reconstruct a regulatory network showing a two-layered regulatory hierarchy. Additional evidence for this hierarchy comes from the analysis of growth phenotypes showing the responses to perturbing members of the upper layer are a superset of the responses to perturbing members of the lower layer. An interpretation would be that the lower layer has a more specialized regulatory than the upper layer. A third data set analyzed shows the transcriptional response of 329 validated RAS pathway target genes to TF perturbations. This part of the analysis is somewhat weaker than the rest. What is the scientific rationale for restricting the analysis to this small set of genes? It's good to have a list of validated targets, but a genome-wide analysis could have shown much more overlap between transcriptional profiles. The observations that patterns of target regulation did not overlap substantially could be completely due to this bias in gene selection. In particular, if the overlap is so small, doesn't that contradict the tight connections we see within the layers in Fig 4? Why are thetwo layers not already visible in Fig 2D, which shows the overlap of transcriptional responses?*

Authors' response: We agree that Fig 2D and its description was misleading, mainly due to confusing data normalization: The figure showed that each transcription factor regulates less than 20% of all RAS pathway target genes, and that two factors typically share 10% of all RAS targets or less. Thus, while the number of shared (co-regulated) RAS targets is *small relative to all RAS targets*, it is *large relative to the number of targets regulated by each factor*. For example, more than half of RelA targets are also regulated by Otx, suggesting strong overlap. Following the reviewer' suggestion, we analyzed target gene expression by interrogating commercial (transcriptome-wide) microarrays and replaced Fig. 2C/D with a more understandable quantification of overlap. These demonstrate a substantial overlap of targets, which were sensitive to transcription factor knockdowns. We had decided to use our customized rat oligonuleotide microarry platform (published in Methods in Enzymology, 2005) initially for the following reasons: The differential

© European Molecular Biology Organization

expression all Ras pathway targets represented on the array had been validated by several *independent* analytical methods (cDNA subtraction, reverse northern analysis, conventional northern analysis or RT-PCR), just trying to respect the "old" tradition in biochemical analysis. We did not consider a potential bias a serious problem, because we focused on transcription factors up-regulated by oncogenic Ras and their targets which are likely to contribute to cellular transformation and malignancy. In spite of the wide coverage of the transcriptome, the commercial rat microarray is of limited value for target gene identification, because the rat genome is less well annotated than the human or mouse genome and the probes often match multiple transcripts. These shortcomings somehow hamper the unambiguous identification of expressed genes. Our conclusions concerning the network organization are now based on the analysis of two independent microarray platforms. As regards the whole transcriptome analysis, most probe sets that correspond to Ras-responsive gene transcripts change significantly upon knockdown of multiple transcription factors, thus providing evidence for a highly connected network (revised Fig. 2D). Moreover, each knockdown affects hybridization of target mRNA to a large number of probe sets on the array, supporting the conclusion that all transcription factors are part of this network (revised Fig. 2C). We agree with the reviewer that the hierarchical patterns should be visible in the gene expression response. We applied several strategies for detecting hierarchies in high-throughput datasets (e.g., nested effect models, gene filtering by expression cut-offs). While each method has its limitations, the conclusion from our analyses was that the gene expression profiles do not show a strict universal hierarchical regulation pattern, which would include all transcriptional regulators and targets in the set analysed. This means that target genes responding to perturbations of downstream factors are not invariably co-regulated by an upstream factor. However, Otx1 knockdown is the perturbation that shows the largest percentage of positive probe sets not responding to any other knockdown. This pattern would be consistent with an upstream role of this transcription factor. To systematically investigate whether target gene expression patterns are consistent with a hierarchical organisation of the regulators, we employed gene filtering using expression cut-offs to identify probe sets whose perturbation-response resembled the patterns of transcription factor mRNA or protein (cf. revised text). As shown in the revised Figs. 3B and C, we found that a large part of the gene expression responses is consistent with the transcription factor expression patterns in Fig. 3A, This finding supports the concept of hierarchical regulation. Specifically, for each of the seven transcription factors, we find genes that match the perturbation response pattern, and may thus be directly regulated by this factor. Taken together, we conclude that the array measurements are partially consistent with hierarchical regulation, but also contain patterns that contradict the hierarchy. At present, it is difficult to understand why these contradictions arise. One explanation may be that hierarchy is less pronounced for target genes, whose transcription is controlled in the same way as the transcription of the seven factors (Fig. 3B). One might further speculate that these genes may be controlled in a combinatorial fashion or that they require input from (upstream) transcription factors not considered in the perturbation response analysis (including transcription factors that are not differentially expressed upon Ras oncogene expression). Following this argument, a perturbation screen containing all 51 RAS regulated transcription factors would be required; we believe that this is beyond the scope of the present work. Therefore, we prefer to rely on double perturbation experiments and phenotypic analyses for supporting the proposed transcription factor hierarchy (see revised manuscript and below). Major changes in the manuscript: - We described the additional whole-genome transcriptome analysis in the text. - In Figure 2, we added panels showing the overlaps in the larger Affymetrix micro-array analysis, and the number of common targets is displayed as histogram. - We modelled the response of the transcriptome and results are shown in Fig. 3.

Reviewer: The rationale behind the growth data is "responses mediated by downstream factors ... were a subset of those mediated by upstream factors". Shouldn't this subset relation also be visible in the expression data? Genes regulated by the downstream factors should also react if the upstream factors are perturbed, while genes directly regulated by the upstream factors should only respond to perturbing the upstream factors. Authors' response: Please see our answer to the previous point.
Reviewer: I don't expect all of this to be visible 100%. I understand about noise, feedback and compensation mechanisms. But the low overlap between transcriptional responses worries me, given that the main result of the paper is a highly connected network.

Authors' response: As stated before, our presentation of the overlap was misleading, as the overlap is relatively high, and we thank the reviewer for pointing this out. The genome-wide expression analysis also confirms that the network is strongly connected. We find that a large number of genes

© European Molecular Biology Organization

are regulated by many of the transcription factors, and our model shows that this is due to the high connectivity within the network.

Reviewer #2 (Remarks to the Author): *In this manuscript, the authors adopt a systems biology approach to dissect the regulatory network involving 7 transcription factors, selected from 51 TFs up-regulated by the KRAS oncoprotein in an in vitro transformation system of rat ovarian epithelial cells. On the basis of a reverse-engineering approach applied to the results generated by perturbation experiments (selective inhibition of signaling pathways and knockdown of each of the 7 TFs followed by RT-PCR and WB analysis), the authors propose a network in which the analyzed transcription factors can be separated in two hierarchically distint groups. In support of the proposed model, the authors show the effect of the members of the two groups of TFs on several growth parameters and target gene expression. Given the relevance of the KRAS-dependent transcriptional networks in human tumorigenesis, this is a very interesting contribution, based on a strategy (MRA, modular response analysis) previously applied to the study of various gene networks. However, the experimental validation of the model prediction is incomplete and not fully convincing.*

Major points:

1. *The inferred regulatory network (Fig.3B and scheme in Fig.4) leads to predictions concerning: a) the reciprocal interactions between between 3 signaling pathways and 7 transcription factors (panels 1,4,7,9 in Fig.3B); b) the reciprocal interactions between the 7 transcription factors (panels 3 in Fig.3B), leading to the distinction between two hierarchical groups (Fig-4). While the proposed transcription factor hierarchy is tested in Fig.5, the model predictions of the effect of signaling pathways on protein levels of TFs are not experimentally verified. This would beparticularly relevant in the case of Fosl1, for which the hitherto unknown model prediction is in striking contrast with multiple independent reports (Vial E. et al., J. Cell Sci. 2003; Casalino et al., MCB 2003; Murphy LO et al., MCB 2004;Basbous et al., 2007) showing that ERK phosphorylation positive controls the level of the Fosl1 gene product (Fra-1). To test the reliability of the proposed model, the authors should show that in their cell system Fra-1 is actually negatively controlled by ERK activity (by investigating, for example, the Fra-1 stability and/or accumulation of the protein ectopically expressed under control of a heterologous promoter, as shown in Basbous et al., 2007).*

Authors' response: We thank the referee for this suggestion. Indeed, the model contradicts previous findings in that it assumes that Fosl1 is destabilized by Erk activity. Therefore, we tested whether Fosl1 is stabilized by Erk also in our cells by expressing Fosl1 ectopically and measuring protein levels after treatment with the MEK inhibitor. Erk does indeed stabilize Fosl1 in our cell system, thus confirming the published reports and invalidating our model (revised Fig. 5 A and B). Nevertheless, our data show that a negative post-transcriptional regulation of the Fosl1 pool is essential to explain the experimental measurements: In Fig. 3A, the Fosl1 mRNA pool is upregulated by several perturbations, while Fosl1 protein is down-regulated at the same time. We confirmed this observation by an independent set of experiments (see revised Fig. 5). In our previous model, this could only be attributed to Fosl1 protein destabilization by pErk, since posttranscriptional interactions between transcription factors were not allowed during the model selection procedure. We speculated that Fosl1 protein stability may also be influenced by other transcription factors (possibly via molecular species not considered in our perturbation screen). Thus, we relaxed our model constraints, and allowed the Fosl1 protein to be post-transcriptionally regulated by other transcription factors in a refined model (cf. revised text). Model selection revealed three potential post-transcriptional interactions controlling Fosl1 protein levels, of which we tested the strongest interaction (Fosl1 protein stabilization by Otx1) by a double-perturbation experiment. These experiments confirm the model prediction, and, interestingly, they point to posttranscriptional regulation in the gene regulatory network without feedback to the MAPK cascade. Major changes in the manuscript: We have added a new Figure in the manuscript that describes these experiments (Figure 5), and added a new section to the manuscript describing it. Notably, the Fosl1 overexpression experiments proposed by the reviewer also allowed us to confirm some aspects of hierarchical organization in the transcription factor network (Fig. 5).

Reviewer: 2.

© European Molecular Biology Organization

Along with the gene expression clusters inferred by expression profiling of the transcripts affected by each of the 7 transcription factors, the authors suggest that the differential effect on 2D growth of RelA/Otx1/Gfi1 vs Hmga2/JunB/Fosl1/Klf6 (Fig.5) reflects the proposed hierarchy between "upstream" and "downstream" transcription factors (Fig.4). For full validation of the model prediction, the authors should test if other parameters of the KRASdependent neoplastic transformation, such as cell motility, invasiveness, anoikis, etc., are differentially affected by the knockdown of members of the two "hierarchically separated" groups of transcription factors.

Authors' response: We performed wound healing assays, which further confirm the hierarchical grouping of transcription factors. The results of these assays are now presented in Figure 6D, E.

Reviewer #3 (Remarks to the Author): *There experimental procedures appear to be valid but the computational analysis does not appear fully cohesive and to the point. There are many small but important issues that need to be address before the paper can be published.*
Specific comments:"So far, signaling kinase inhibitors which target the PI-3-K/Akt- or Raf/Mek/Erk-pathways have shown limited therapeutic efficacy in the clinical setting." Please elaborate with references.

Authors' response:
Two references have been inserted (page 3). In view of still ongoing experimental and clinical trials using MAPK and PI3K inhibitors and ongoing mechanistic studies, a large number of references should be quoted to cover the subject properly. However, we believe that this would be beyond the scope of this paper.

Reviewer: "Microarray analysis permits the simultaneous assessment of many network elements. However, the derivation of the network topology from microarray data is not straight-forward, as network perturbation data are lacking". Most microarray analyses are done on perturbation data. The problem with microarrays is that they measure mRNA levels which do not directly reflect regulatory mechanisms. This is not clearly stated by the authors.

Authors' response: We have changed the text accordingly (page 3)

Reviewer: "Bayesian statistics, neural network type ODE models (Busch et al, 2009) and metabolic control analysis (Kholodenko et al, 2002)" More examples are needed to show appropriate coverage of the topic. Tools such as ARACNE the infderelator and eureqa should be mentioned.

Authors' response: We have expanded the section on page 4 and quoted these and other additional methods.

Reviewer: Page 4 "contrasted" I think it was meant "constructed"

Here we meant "contrasted" (and previously published it), because subtractive hybridization and the performed microarray analysis have compared two different conditions (normal versus RAStransformed). We have changed contrasted to compared, to avoid confusion.

Reviewer: "the adhesion molecule E-cadherin (Malek et al., unpublished data)." If microarrays were performed and the data was previously published why there is a reference to Malek et al., unpublished data?

Authors' response: These unpublished data is not necessary for the manuscript and the reference was removed from the introduction.

Reviewer: Explain "anoikis" for the non-expert reader.

Authors' response: We thank the reviewer for pointing this out. Anoikis designates programmed cell death induced by depriving anchorage dependent cells (such as "normal" epithelial cells) from their substratum (e.g. the plastic cell culture dish). We deleted this from the introductory text because the term anchorage independence is sufficient to describe the phenotypic properties in KRAS transformed cells.

© European Molecular Biology Organization

Reviewer: Page 4 "Reversely" should be "reverse"
Authors' response: We changed this.

Reviewer: Page 4 "focussed" should be "focused"
Authors' response: Both spellings are OK in British English according to Macmillan Dictionary.

Reviewer: "we focussed on the transcription factors Fosl1, Gfi1, Hmga2, JunB, Klf6, and Otx1" Is it known that these factors regulate the genes that changed in expression? Do the factors have binding motif sequences at the promoters of the genes that changed in expression? Is there ChIP evidence for any of these factors? Can the kinase be linked to the factors through known pathways?

Authors' response: The selection process was as follows: We found 51 transcription factors that showed up-regulation. We included two of them, Hmga2 and Fosl1 since they are known to contribute to malignant phenotype. We included four additional factors (Otx1, Klf6, Gfi1, JunB), since they met several functional criteria (high expression in inflammatory responses, tissue damage, tumourigenisis and repression in differentiated cells). In addition, we included RelA as a transcription factor that is regulated on the post-translational level but is known to contribute to the malignant phenotype. This is now described in detail in the introduction.

Reviewer: "RelA was not included in the KRAS pathway-dependent gene signature, but was upregulated at the protein level" protein level experiments never introduced before this comment. The reader does know about these experiments yet.

Authors' response: We have changed this section and cited the relevant paper, in which RelA was shown to be post-translationally regulated by Ras.

Reviewer: "modular response analysis" approach should be explained at the bottom of page 4 and not so much in the results.

Authors' response: We moved this part to the introduction.

Reviewer: Most of page 6 belongs in the introduction.
Authors' response: We moved the description of transcription factors selection to the introduction.

Reviewer: Direct physical interaction between RelA and Hmga2 should be mentioned (PMID: 12693954).
Authors' response: We added this reference together with other references of previously known interactions to the supplement.

Reviewer: "customised microarrays representing 329 independently validated RAS pathway-target genes. KRAS oncogene-mediated transformation of ROSE cells was associated with the deregulation of 147 target genes." Not clear if the 147 are a subset of the 329? If the 147 are a subset of the 329, why the rest of the genes are on the microarray? Need clear and detailed explanation.

Authors' response: Our group has been working on the identification and characterization of RAStarget genes for almost 20 years (cf. Zuber et al., Nature Genetics, 2000), and the 329 genes represent target that have been identified in diverse experimental systems such as fibroblasts, epithelial cells and carcinomas. Thus, in any given cell type, only a subset of these genes are expressed and regulated by RAS. As we have now performed transcriptome-wide measurements, we have shortened the descriptions of the results from the customized arrays, and described the selection procedure of the set of genes for the customized arrays in detail in the materials and methods part (page 31).

Reviewer: "We concluded that the RAS pathway orchestrated the de-regulation of the transcriptome via a limited number of transcription factors, each of which controls the transcription of distinct subsets of target genes." This conclusion is problematic as there could be many other explanations for the observed results. Randomized controls need to be created. Also, there is no mention of genes that were aggravated by the knockdown, i.e. further increased or further decrease in expression due to the knock-down.

© European Molecular Biology Organization

Authors' response: Our previous presentation of these results was misleading, mainly due to confusing normalization of the data (see response to reviewer 1): The previous Fig. 2D showed that each factor regulates less than 20% of all RAS-regulated genes, and that two factors share typically 10% or less of all RAS targets. Thus, while the number of shared RAS targets is *small relative to all RAS targets*, it is *large relative to the number of targets regulated by each factor*. For example, more than 50% of RelA targets are also regulated by Otx, thus suggesting strong overlap. In the revised manuscript, we replaced the previous Fig. 2C/D with an analysis of a genome-wide dataset (see below), which shows even more substantial overlap between knockdowns. Thus, the array data is fully consistent with a highly interconnected network of transcription factors. In a randomized control, the probability of a gene being regulated by multiple transcription factor would be much smaller than found here (revised Figure 2D). There were roughly as many genes regulated by multiple (or even 7) knockdowns as by a single knockdown. Thus, given that our scrambled siRNAs affected few genes only, it is highly unlikely that these results can be explained by chance only. We agree with the referee that the transcriptome analysis alone does not allow the conclusion that the set of transcription factors investigated here each controls a distinct subset , and therefore removed this sentence from the manuscript.

Reviewer: Figure 3A is most central to the paper but difficult to understand. From the figure it looks like inhibiting pErk does not have an effect or has an effect that is similar to the over-expression of Ras. Is this true? It is also difficult to understand how the MRA algorithm came to the conclusions it did given the results from 3A. The colors do not match.

Authors' response: Figure 3A showed values proportional to the global response coefficients. These are defined in such a way that they are positive, if the response follows the same direction as the perturbation. As RAS expression positively affects most of the transcription factors, and Mek inhibition negatively regulates them, both effects have positive response coefficients. We agree that this is difficult to understand, and thus plotted log2 fold-changes in the revised manuscript (which are the same numerical values as in the previous Figure, but have in part opposing signs for all inhibitor- and knockdown experiments) in Figure 3A. In response to the comment below, we also present the data as a heat map.

Reviewer: "calculated the correlation coefficients between transcription factor and target gene mRNAs over 9 experimental perturbation conditions" Not clear why this was done this way and not just compare the expression of the genes fold change or just change vs. the control. There is not mention in the methods how the correlation coefficients were calculated.

Authors' response: The idea was that direct targets of a transcription factor should show the same perturbation-response pattern as the transcription factor itself (i.e., the perturbation-response patterns should have a high correlation coefficient). We agree with the reviewer that the previous presentation needs improvement; we therefore replaced the correlation analysis by a figure that directly compares perturbation-response patterns of transcription factors and target genes in the genome-wide arrays (revised Fig. 3B and C). In the construction of this figure, we used gene filtering using expression cutoffs which turned out to be more robust against artifacts when compared to correlation analysis. Figs. 3B and C show that many gene patterns are closely related to transcription factor patterns, thus suggesting that these are indeed direct target genes of the corresponding transcription factors.

Reviewer: The way the global response matrix R depicted in Figure 3A is too complex and not producing clear results. It should be replaced with a simpler procedure.

Authors' response: See response before. In the revised ms. we do have plotted the fold-changes rather than the response coefficients, relabeled the columns, and used a heat map to illustrate the results.

Reviewer: "feedback regulation of exogenously overexpressed RAS protein by other nodes in the network" Not clear. Why not shown?

Authors' response: Since the KRAS-transformed cell line has been characterized in a previous publication, we did not show the difference between endogenous, wild-type RAS levels and

© European Molecular Biology Organization

ectopically expressed, oncogenic RAS, which is much more abundant than wild-type RAS (Tchernitsa et al., Oncogene, 2004). Since oncogenic RAS is constitutively active (i.e., always bound to GTP), we can assume that the total amount of active RAS-GTP correlates well with the total expression of oncogenic RAS. Since the expression of oncogenic RAS is not affected by the knockdowns, we can neglect feedback regulation upstream of RAS. In the revised Fig. 3A, these measurements would correspond to an empty (white) row, therefore we decided to omit this (note: Fig. 3A contains a RAS perturbation column but not a RAS expression row).

Reviewer: "The best model fit agrees well with the measured perturbation response data (Fig. 3C)." This needs to be further quantified and explained.

Authors' response: In the revised ms,. we report the Chi2 value and show the experimental variability in the scatter plot comparing experimental data and the model (previously Fig 3C, now Fig. 4C).

Reviewer: The reduced growth by knocking down the transcription factors is not a valid validation of the model by itself since such experiments could have been done regardless of the modeling.

Authors' response: We agree that the hierarchical model primarily helps to understand these experimental results, but was not instructive for the experimental design. We made this now clear in the results. In order to provide direct testing of a model prediction, we employed the Fosl1 overexpression system, as proposed by reviewer 2, to confirm the model more directly by modeldriven experiments. Specifically, we investigated post-transcriptional interactions among transcription factors and confirmed model-predicted hierarchical interactions between Fosl1 and Hmga2 (revised Fig. 5).

Reviewer: The MRA model is relatively complex so one wonders if the separation between Otx1, Gfi1 and RelA vs. Hmga2, Fosl1, Klf6 and JunB that was shown for the validation in Fig. 5 could have been identified by correlation analysis.

Authors' response: The problem with correlation analysis is that it does not help to identify directionality. In the particular network described, we recognize a clear directionality from the upper to the lower level. For example, there is a strong directional link from Otx to Fosl, so these correlate strongly. However, correlation will not assign Otx to the upstream part of the network and Fosl1 to the downstream part. In this regard, MRA is clearly superior to correlation analysis.

Reviewer: What is the relationship between the transcription factors within each layer?

Authors' response: The transcription factors within each layer may also show hierarchical regulation or they may act relatively independent of each other. The lower layer appears to be organized hierarchically: neglecting feedbacks via the signaling network, we observe a transcriptional cascade (KLF6 -> Fosl1 -> JunB -> Hmga2), whose directionality is intercepted by a single interaction only (JunB -> Klf6). Accordingly, Fosl1 and Klf6 knockdowns control more factors of the lower layer at the protein level when compared to JunB and Hmga2. In the revised Fig. 5, we confirm some aspects of this unidirectional cascade using double perturbation experiments. The factors of the upper layer appear to act relatively independent of each other: knockdown of each factor in the upper layer hardly affects the other factors in the same layer (Fig. 3A). This suggests that the corresponding MRA interactions are relatively weak. Accordingly, we observe that many target genes are selectively controlled by one of the factors in the upper layer

Reviewer: In figure 5E legend and text it says 7 + 2 = 9 conditions but only 7 are shown, what about the Akt and Erk conditions? How come these are not a part of the clustergram?

Authors' response: The former Fig 5E showed correlation with the expression of the 7 transcription factors over all perturbation conditions. We removed this figure from the paper, and replaced it by the revised Figs. 3 and C (see response above). In the revised Figs. 3 B and C, we indeed show 7 seven knockdown conditions only; this is due to the fact that genome-wide transcription profiles were measured only for these 7 conditions to investigate interactions among transcription factors. In summary, we do hope that we have been able to answer all reviewers' queries to their satisfaction

© European Molecular Biology Organization

and have provided sufficient new information. We are very much looking forward to your editorial decision on our new manuscript.

2nd Editorial Decision 23 May 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 the study. As you will see, the referees find the topic of your study of potential interest and are largely supportive. They raise however a series of important concerns and make suggestions for modifications, which we would ask you to carefully address in a revision of the present work.

Reviewers #2 and #3 both reviewed the previous version of this work (MSB-12-3730), they generally agreed that the revisions and new experimental evidence had substantially improved this work (Reviewer #2 was #1, and #3 was #2 previously). The reviewers, however, felt that additional effort was needed to explain the regulatory model construction and to quantify the robustness of the model edges (in this regard the first major points from reviewers #1 and #2 seem very closely related).

Addressing these concerns, and the other points raised by the reviewers may require additional analyses, and we reserve the right to send any revised work back to some or all of the reviewers if needed.

When preparing your revised work, please also address the following issues:

Molecular Systems Biology generally requires that all key experimental data are made available with research publications.

- Please deposit the expression array data in a public repository like GEO, and include the resulting accession numbers in the Methods section. This should also include the custom RAS-targeted microarray data -- if these data cannot be included in a public repository for some reason, they can also be provided as supplementary materials.

- In addition to our capacity to host datasets in our supplementary information section, we provide a new functionality that allows readers to directly download the 'source data' associated with selected figure panels (e.g. <http://tinyurl.com/365zpej>). This sort of figure-associated data may be particularly appropriate for Figures 2, 5, and 6. Please see our Instructions of Authors for more details on preparation and formatting of figure source data (<http://www.nature.com/msb/authors/index.html#a3.4.3>). This same system can also be used to supply replicate gel images, for example for Fig 1.

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):

Schafer and colleagues describe a study that aimed at elucidating a KRAS pathway -dependent transcriptional network. Receptor Tyrosine Kinase regulated signaling is critical in a number of mechanisms during tumourigenesis. The authors systematically perturbed the KRAS signaling pathway by transfecting siRNAs targeting seven out of 51 transcription factors (TFs) that they had previously identified to be up-regulated in rat ROSE cells upon KRAS-transformation. These TFs

© European Molecular Biology Organization

were selected based on their known functional involvement in signaling as well as tumor-related phenotypes. RelA was selected on top of six RTK-related TFs, as NF-KB-signaling is known to also impact on RAS-driven transformation. Remark: it will be interesting to follow-up on RelA and p65 also because only a minor fraction of the cytosolic p65 protein translocates to the nucleus upon activation while the majority of p65 is kept cytoplasmic thus not contributing to NF-KB signaling. Hence, changes even in p65 abundance do not directly translate into elevated NF-KB signaling. Because of the highly dynamic nature of NF-KB signaling this would be required to test at single cell level via microscopy methods.

The authors performed knockdown experiments for each of the seven TFs and then carried out expression profiling analysis to detect regulated genes. Indeed they found a number of genes that were deregulated upon knockdown of the individual transcription factors. Interestingly, the majority of these differentially expressed genes were deregulated by two or more of the seven transcription factors under investigation. This led the authors to conclude that KRAS-signaling would be regulated by a dense and connected signaling network. This conclusion could suggest that perturbations of the individual TFs might regulate some of the other selected TFs. Consequently, the abundance of all seven TF proteins was investigated in response to knockdowns by Western Blot. And indeed, up- or down-regulation of the majority of the seven TFs was observed at RNA and/or protein levels.

Since the experimental design had not been restricted to identifying directly regulated genes but could as well detect secondary events, the authors next tested their gene expression datasets for such genes that showed similar expression patterns as the selected TFs were affected by knockdown of either of the TFs at protein (Figure 3B) and RNA (Figure 3C) levels. This identified a number of 947 probe sets that indeed displayed transcription patterns that were similar to the protein patterns of the individual TFs. This was indication of a direct regulation of these genes by the tested TFs. Reverse engineering of the network topology then helped to distinguish direct from indirect effects and provided a hierarchical organization of the transcription factors under investigation. While Gfi1, Otx1 and RelA should be located upstream in the network, the other TFs appeared to be acting further downstream not feeding back on Gfi1, Otx1 or RelA. An unexpected prediction of their model involved Fosl1 and Erk signaling, as the model suggested the TF to be destabilized by active Erk-signaling. (Remark: Erk signaling is commonly elevated temporarily upon addition of ligands or the application of stresses. How stable is the activation state in the steady state conditions of the experiments that are described?) This predicted effect of Erk-signaling on Fosl1 levels was in contrast to the literature, hence the authors tested their model in a Fosl1 overexpression system after addition of the U0126 Mek inhibitor. The resulting data demonstrated that the initial model was wrong in at least this interaction, and the model was consequently refined to take reference to the new data. The revised model was then further tested as the authors performed experiments interrogating phenotypes induced by the different perturbations (cell cycle phase, anchorage-dependent and -independent growth, collective cell migration).

The authors had started from a larger number of TFs that are deregulated in a KRAS-transformed cell line model. They selected seven of the TFs and, driven by experimental data, generated a model that could predict the behavior of the individual TFs within a TF-interaction network. Most of the interactions had been known already before (literature). It would be interesting to see how many known interactions were not predicted/observed and could thus be false-negative interactions in the model (compare Figure 4A). Furthermore, in the future the authors might try to confirm some of the interactions that are predicted in Figure 4A panels 3 and 9.

Major points
Why does the model shown in Figure 4C not show edges between pAkt and RelA, Otx1 or Hmga2? The data in Figure 1 and also the color code in Figure 3A suggests a strong regulation of Otx1 and Hmga2 by PI3K/Akt-signaling. This is somewhat contradictory to statements in the first paragraph of the Results part where regulation of Fosl1, Hmga2, Otx1, Klf6, Gfi1 and JunB is said to be strictly dependent on Raf/Mek/Erk signaling. Would acceptance of a coregulation of Otx1 and Hmga2 by Akt and MAPK signaling affect the model predictions?

The legend for Figure 1 does not give the number of replicates that were generated. In the legend for Supplementary Fig. 1 (also in Supplementary Figure 4) the authors should state whether the indicated 'two independent experiments' were technical or biological replicate experiments. Ideally,

© European Molecular Biology Organization

two biological replicates with two technical replicates each should be carried out, since the reproducibility of quantitative Western Blotting is rather poor. How well did the individual TFs reproduce in the two experiments being mentioned? Since the entire model is based on this protein data a better description on the relibility of that data should be provided.

Minor point
Even if the rat Affymetrix arrays are not well annotated a mapping of probe sets onto genes should be done in order to test the reproducibility of results from two or more probe sets representing the same genes and to obtain a better idea on the numbers of genes that are up- or down-regulated in response to the perturbation experiments (in all, 7561 probe sets are said to represent 1757 genes). The numbers and identities of genes should be more relevant than probe numbers in the following, e.g., when the numbers of probe sets are given indicating up- or downregulation after TF knockdown. For example, the authors should make clear on page 7 when they speak about '1,081 (18%) of the transcription factor targets' when they refer to Figure 2D whether they mean probes or genes (I would assume they mean probe sets).

Reviewer #2 (Remarks to the Author):

This paper has significantly improved compared to its first submission. My main concern had been lack of overlap in the transcriptional patterns observed after perturbing TFs in the two network layers and a potential bias in only using at RAS-specific microarray. These concerns have been resolved.

In my following comments I will concentate on the authors' method and its presentation:

It would strengthen the paper if the strength of evidence for each of the inferred edges could be quantified. Right now the graph in 4C is the model with the hightest fit. Is there a way to use cross-validation or bootstrapping to test how robust this model is? Correlation of predictions with left-out data would be a stronger evidence for the models power than correlation with the data used for fitting (Fig4B).

The description of model construction and evaluation is not clear to me. In the key equation ((r-1)s - pR)2, what is s? what is p? If p is a vector, then pR should be a vector too, are you doing separate optimizations for rows of r^-1? I also do not understand the relationship between the Levenberg-Marquardt algorithm (LMA) used to fit r and p, and the step wise approach desribed directly after that. Both optimization approaches lead to local, not necessarily global optima. Did the authors check different initializations for LMA? Would the step-wise procedure lead to a similar model as the one reported if it started from a random r matrix and added/deleted singel edges to improve fit?

My goal is not to keep the authors busy with more and more simulations and changes to the algorithm. The paper looks very mature as it is. But the model construction is at the heart of this paper and I see a need to explain it better, including robustness and possible limitations.

Reviewer #3 (Remarks to the Author):

This very interesting manuscript has been substantially improved in its revised version. I am particularly impressed by the number of additional microarray assays and functional assays which have been included.
The authors have fully responded to all the points raised in the previous revision. With regard to the unexpected posttranscriptional regulation of fosl1, the authors have refined the proposed model, on the basis of the novel experimental data now shown in figure 5.
In addition, the authors have further validated the proposed hierarchical grouping of the two sets of transcription factors by cell migration along with 2D growth assays, as shown in figure 6.

© European Molecular Biology Organization

1st Revision - authors' response 26 June 2012

Thank you very much for your interest in publishing our work. Please find below a point-by-point reply to your suggestions and the reviewers' comments.

Editor: Please deposit the expression array data in a public repository like GEO, and include the resulting accession numbers in the Methods section.

Authors' reply: Affymetrix arrays: GSE 38584, GSE 38585 (page 33); customized arrays: GSE 24668 (page 32)

Editor: In addition to our capacity to host datasets in our supplementary information section, we provide a new functionality that allows readers to directly download the 'source data' associated with selected figure panels (e.g. <http://tinyurl.com/365zpej>). This sort of figure-associated data may be particularly appropriate for Figures 2, 5, and 6. This same system can also be used to supply replicate gel images, for example for Fig 1.

Authors' reply: DONE

Reviewer #1 (Remarks to the Author):

Schafer and colleagues describe a study that aimed at elucidating a KRAS pathway -dependent transcriptional network. Receptor Tyrosine Kinase regulated signaling is critical in a number of mechanisms during tumourigenesis. The authors systematically perturbed the KRAS signaling pathway by transfecting siRNAs targeting seven out of 51 transcription factors (TFs) that they had previously identified to be up-regulated in rat ROSE cells upon KRAS-transformation. These TFs were selected based on their known functional involvement in signaling as well as tumor-related phenotypes. RelA was selected on top of six RTK-related TFs, as NF-KB-signaling is known to also impact on RAS-driven transformation.

Remark: it will be interesting to follow-up on RelA and p65 also because only a minor fraction of the cytosolic p65 protein translocates to the nucleus upon activation while the majority of p65 is kept cytoplasmic thus not contributing to NF-KB signaling. Hence, changes even in p65 abundance do not directly translate into elevated NF-KB signaling. Because of the highly dynamic nature of NF-KB signaling this would be required to test at single cell level via microscopy methods.

Authors' reply:
We fully agree with the reviewer in that the pertubations affecting mRNA and protein levels are quite crude and do not allow conclusions about the dynamics of NF-KB signaling (cytosolic versus nuclear effects) and e.g. about the complex interactions of the components of the AP-1 transcription factor complex (FOS/FOSL1/JUN). The network analysis described in our ms. allows a basic reconstruction of the interactions of RAS pathway-modulated, up-regulated transcription factors and their contribution to phenotypic read-outs. This was our main objective. Doing the analysis at the single cell level is an excellent suggestion for future work. We would envisage using Panomics *in situ* hybridization technology for detecting low-abundance mRNA in single cells. For protein analysis, immunohistochemical or immunofluorescence analysis would have to be established for all factors.

The authors performed knockdown experiments for each of the seven TFs and then carried out expression profiling analysis to detect regulated genes. Indeed they found a number of genes that were deregulated upon knockdown of the individual transcription factors. Interestingly, the majority of these differentially expressed genes were deregulated by two or more of the seven transcription factors under investigation. This led the authors to conclude that KRAS-signaling would be regulated by a dense and connected signaling network. This conclusion could suggest that perturbations of the individual TFs might regulate some of the other selected TFs. Consequently, the abundance of all seven TF proteins was investigated in response to knockdowns by Western Blot.

© European Molecular Biology Organization

And indeed, up- or down-regulation of the majority of the seven TFs was observed at RNA and/or protein levels.

Since the experimental design had not been restricted to identifying directly regulated genes but could as well detect secondary events, the authors next tested their gene expression datasets for such genes that showed similar expression patterns as the selected TFs were affected by knockdown of either of the TFs at protein (Figure 3B) and RNA (Figure 3C) levels. This identified a number of 947 probe sets that indeed displayed transcription patterns that were similar to the protein patterns of the individual TFs. This was indication of a direct regulation of these genes by the tested TFs.

Reverse engineering of the network topology then helped to distinguish direct from indirect effects and provided a hierarchical organization of the transcription factors under investigation. While Gfi1, Otx1 and RelA should be located upstream in the network, the other TFs appeared to be acting further downstream not feeding back on Gfi1, Otx1 or RelA. An unexpected prediction of their model involved Fosl1 and Erk signaling, as the model suggested the TF to be destabilized by active Erk-signaling. (Remark: Erk signaling is commonly elevated temporarily upon addition of ligands or the application of stresses. How stable is the activation state in the steady state conditions of the experiments that are described?) This predicted effect of Erk-signaling on Fosl1 levels was in contrast to the literature, hence the authors tested their model in a Fosl1 overexpression system after addition of the U0126 Mek inhibitor. The resulting data demonstrated that the initial model was wrong in at least this interaction, and the model was consequently refined to take reference to the new data. The revised model was then further tested as the authors performed experiments interrogating phenotypes induced by the different perturbations (cell cycle phase, anchorage-dependent and -independent growth, collective cell migration).

Authors' reply to remark on the stability of Erk activation:

KRAS-transformed ROSE cells are stable transfectants. In general, normal precursor cells (e.g. NIH/3T3 cells) contain 500 fmol GDP-RAS and 1.3 fmol GTP-RAS. Levels rise up to 5000 fmol GDP-RAS and 2000 fmol GTP-RAS, respectively, in cells expressing mutant RAS due to multiple integrations of the transgene (cf. Robin Hesketh, The oncogene and tumor suppressor gene facts book, Academic Press 1997, p. 319ff). Hence, a strong and sustained Erk activation is achieved over many cell generations. In addition, we observed stable Erk activation in cells expressing mutant RAS controlled by an inducible promoter (cf. Lund et al., Oncogene 10, 4890-903 [2007]).

The authors had started from a larger number of TFs that are deregulated in a KRAS-transformed cell line model. They selected seven of the TFs and, driven by experimental data, generated a model that could predict the behavior of the individual TFs within a TF-interaction network. Most of the interactions had been known already before (literature). It would be interesting to see how many known interactions were not predicted/observed and could thus be false-negative interactions in the model (compare Figure 4A). Furthermore, in the future the authors might try to confirm some of the interactions that are predicted in Figure 4A panels 3 and 9.

Authors' reply:
We agree with the reviewer that a fair assessment of a reverse engineering algorithm should also address the false-negative rate. Therefore, we performed a systematic analysis addressing the reliability of our reverse engineering algorithm using benchmark models that have a similar size as the real data set (see also our response to Reviewer 2). The algorithm has a very high precision (> 80%), but showed a considerably higher number of false-negatives. This is, in part, due to the fact that weak (and, hence, less important) interactions are removed from the model. Thus, even if an interaction exists in the network, it will not be ranked highly by the algorithm, if it does not strongly affect the fitting result. Another problem is that interactions reported in the literature are often specific to certain cell types / species. These interactions cannot *a priori* be assumed to be valid in the rat epithelial system we used in our work. For example, intensive crosstalk has been reported to occur between Akt and Erk signaling cascades, and the interactions were either positive, negative or absent, depending on the cell type and stimulation conditions. Therefore, the finding that our experiment and model failed to detect crosstalk between Akt and Erk does not contradict the existing reports in the literature.

© European Molecular Biology Organization

A more detailed analysis of panels 3 and 9 (Fig. 4A) would indeed be of great interest for future studies. Panel 9 represents potential transcriptional feedback loops. In the current study, an in-depth analysis of such feedback was hindered by the poor annotation of rat microarrays. Thus, we could not easily derive hypotheses on the mechanisms by which the transcription factors affected Akt/Erk signaling. A detailed qPCR-based analysis of known transcriptional feedback regulators would be needed here. Panel 3 displays transcriptional interactions between transcription factors, and thus directly refers to the core transcription factor network. To provide proof-of-principle for demonstrating the power of an integrated functional and computational analysis of the transcription factor network downstream of the oncogenic RAS/MAPK signaling cascade, we restricted our analyses to seven transcription factors. It is indeed possible that some of the direct interactions (in the model) are in fact indirect in the cell (i.e. are mediated by transcription factors not considered in the model). It will be highly desirable to expand the model to all up-regulated transcription factors also considering chromatin-based (e.g. chromatin immunoprecipitation) and promoter assays.

Major points
Why does the model shown in Figure 4C not show edges between pAkt and RelA, Otx1 or Hmga2? The data in Figure 1 and also the color code in Figure 3A suggests a strong regulation of Otx1 and Hmga2 by PI3K/Akt-signaling. This is somewhat contradictory to statements in the first paragraph of the Results part where regulation of Fosl1, Hmga2, Otx1, Klf6, Gfi1 and JunB is said to be strictly dependent on Raf/Mek/Erk signaling. Would acceptance of a coregulation of Otx1 and Hmga2 by Akt and MAPK signaling affect the model predictions?

Authors' reply:
Regulation of RelA, Otx1 and Hmga2 by pAkt is necessary to match the experimental observations: PI3K inhibition affected RelA and Hmga2 proteins, while the corresponding mRNAs remained essentially unaffected (Fig. 1). The mRNA and protein levels of Otx1 coordinately responded to PI3K inhibition. The model explains these effects by post-transcriptional regulation of RelA protein via pAkt. The regulation of the RelA protein by pAkt is indicated in Fig. 4A (box 4). Moreover, it is also visible in the scheme displayed in Fig. 4C (see the yellow box at the top-left of RelA protein). RelA in turn regulates the expression of Otx1 and Hmga2 in the model. In this way, pAkt can affect the levels of RelA protein and of its immediate neighbors Otx and Hmga2 in the model. In conclusion, the model shows the expected edges between pAkt and RelA, Otx1 or Hmga2. Therefore, the existing model predictions remain valid.

The reviewer criticized that pAkt-mediated regulation of Otx1 and Hmga2 "is somewhat contradictory to statements in the first paragraph of the Results part, where regulation of Fosl1, Hmga2, Otx1, Klf6, Gfi1 and JunB is said to be strictly dependent on Raf/Mek/Erk signaling". Here we mainly referred to regulation at the mRNA level. We still think that this claim holds true, but rephrased the corresponding paragraph on page 6 of the manuscript, distinguishing more carefully between regulation at the mRNA and protein levels.

The legend for Figure 1 does not give the number of replicates that were generated. In the legend for Supplementary Fig. 1 (also in Supplementary Figure 4) the authors should state whether the indicated 'two independent experiments' were technical or biological replicate experiments. Ideally, two biological replicates with two technical replicates each should be carried out, since the reproducibility of quantitative Western Blotting is rather poor. How well did the individual TFs reproduce in the two experiments being mentioned? Since the entire model is based on this protein data a better description on the relibility of that data should be provided.

Authors' reply: The data represent 2 biological replicates. Additional data have been up-loaded onto the site provided by MSB.

Minor point
Even if the rat Affymetrix arrays are not well annotated a mapping of probe sets onto genes should be done in order to test the reproducibility of results from two or more probe sets representing the same genes and to obtain a better idea on the numbers of genes that are up- or down-regulated in response to the perturbation experiments (in all, 7561 probe sets are said to represent 1757 genes). The numbers and identities of genes should be more relevant than probe numbers in the following, e.g., when the numbers of probe sets are given indicating up- or downregulation after TF

© European Molecular Biology Organization

knockdown. For example, the authors should make clear on page 7 when they speak about '1,081 (18%) of the transcription factor targets' when they refer to Figure 2D whether they mean probes or genes (I would assume they mean probe sets).

Authors' reply:
In the revised manuscript we clarified that the numbers refer to probe sets, and included the number of genes that are represented by these probe sets in brackets, where appropriate. To estimate the number of target genes more reliably, we now used the xmapcore data base, which is known to provide an improved annotation superior to the annotation files provided by the microarray manufacturer.
To assess reproducibility, we checked how often probe sets of one gene are assigned to one particular hybridization pattern. For the majority of genes represented by multiple probe sets, all regulated probe sets followed the same pattern. We added this finding to the results section (c.f. pages 7 and 8 for clarification of the relationship between probe sets and genes).

Reviewer #2 (Remarks to the Author):

This paper has significantly improved compared to its first submission. My main concern had been lack of overlap in the transcriptional patterns observed after perturbing TFs in the two network layers and a potential bias in only using at RAS-specific microarray. These concerns have been resolved. In my following comments I will concentrate on the authors' method and its presentation: It would strengthen the paper if the strength of evidence for each of the inferred edges could be quantified. Right now the graph in 4C is the model with the highest fit. Is there a way to use cross-validation or bootstrapping to test how robust this model is? Correlation of predictions with left-out data would be a stronger evidence for the models power than correlation with the data used for fitting (Fig4B).

Authors' reply:
Since we use a likelihood-based framework, we can apply the likelihood ratio test to quantify the evidence of each link. Therefore, we systematically removed each link and calculated the difference in likelihood, from which we derive a p-value for each edge. Cross-validation is not possible, since the model becomes non-identifiable on incomplete data sets, and the model parameters cannot be fitted. For example, the specific behavior of the transcription factor mRNA and protein species cannot be correctly inferred when only the mRNA measurements were used for fitting. Depending on the structure of the network, similar problems may even emerge when the mRNA measurements and a subset of the protein measurements are used for fitting.

Before applying our method to the data, we characterized the robustness of our method by model fitting on simulated data of similar size as the real perturbation matrix. We found that the precision was above 80% in the simulated networks. We also found that the false negative rate was substantially higher, such that our approach discovered mainly strong links and rarely detected weak links. Thus, the algorithm is relatively conservative. We also inspected the influence of noise and perturbation strength. Interestingly, the false positive rate is not sensitive, however, the sensitivity of the method increases if the data is less noisy, and if the perturbations are stronger. We added this information to the supplementary text and figures, and also added a summary of this analysis to the main text (page 11).

The description of model construction and evaluation is not clear to me. In the key equation ((r-1)s - pR)2, what is s? what is p? If p is a vector, then pR should be a vector too, are you doing separate optimizations for rows of r^-1? I also do not understand the relationship between the Levenberg-Marquardt algorithm (LMA) used to fit r and p, and the step wise approach desribed directly after that. Both optimization approaches lead to local, not necessarily global optima. Did the authors check different initializations for LMA? Would the step-wise procedure lead to a similar model as the one reported if it started from a random r matrix and added/deleted singel edges to improve fit?

Authors's reply:
We fully agree that our algorithm may lead to local optima, because we combine a local parameter estimation algorithm (LMA) with greedy hill climbing (which scans the possible model topologies

using a local strategy). We initially tried to screen possible model topologies using randomly initialized network structure matrices. However, it is computationally not feasible to reliably find good fits in the space of possible model topologies.

The parameter estimation for each model topology was done by LMA. We tested whether random initialization of the new parameter changed the result after an edge has been added. We found that the algorithm is robust towards initialisation of the parameter for the added link. We fully agree with the reviewer that a global optimization strategy, or strategies such as MCMC may improve the method to better explore the structure and ensure that the global optimum is reached. This will be especially important for larger studies, and MCMC could be also used to quantify the evidence for each edge. We have revised the main text and Materials & Methods describing the algorithm (pages 33 and 34), and added a more detailed description to the supplement.

My goal is not to keep the authors busy with more and more simulations and changes to the algorithm. The paper looks very mature as it is. But the model construction is at the heart of this paper and I see a need to explain it better, including robustness and possible limitations.

Authors's reply:
In the revised ms. we describe the robustness and limitation of the method due to its local optimization strategy (pages 10, 33 and 34).

Reviewer #3 (Remarks to the Author):

This very interesting manuscript has been substantially improved in its revised version. I am particularly impressed by the number of additional microarray assays and functional assays which have been included.
The authors have fully responded to all the points raised in the previous revision. With regard to the unexpected posttranscriptional regulation of fosl1, the authors have refined the proposed model, on the basis of the novel experimental data now shown in figure 5.
In addition, the authors have further validated the proposed hierarchical grouping of the two sets of transcription factors by cell migration along with 2D growth assays, as shown in figure 6.

Authors' reply: We couldn't agree more.

Corrigendum:
When cross-checking the figures we realized that we had mistakenly presented the data related to the Fosl1 over-expression experiment (Fig. 5F). We corrected the figure accordingly. The correct data show that the rescue of Erk-dependent Hmga2 expression by Fosl1 over-expression is stronger than predicted by the model. The conclusion is that Fosl1 by itself has a negligible effect on Hmga2.

We do hope that the current revision will now satisfy all reviewers' comments and are very much looking forward to your final editorial decision.

Acceptance letter 29 June 2012

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, in principle, for publication.

Before we can send this work to production we have some relatively minor remaining issues and requests for modifications:

1. In Fig. 1B, are the control GADPH lanes a repetition of the same gel image (they look very similar)? Experimentally, I think this would be entirely acceptable, but if they are the same repeated image this should be explained clearly in the figure legend. Indeed, if they are the same, I encourage

© European Molecular Biology Organization

you to only show this control gel once to avoid any possible reader confusion, even if it makes the figure somewhat less symmetrical between the RT-PCR and the Western results.

2. Regarding Fig. 5 & 6, journal policy discourages the display of error bars or standard deviations when there are only two independent replicates. I would suggest that you consider just plotting both values individually next to or on top of the bars.

3. The current title is somewhat longer than our format requirements allow (max 100 characters, including spaces). The editor would like to suggest, "Reverse-engineering a hierarchical regulatory network downstream of oncogenic KRAS." This is only a suggestion, please feel free to select a title that you feel best represents this work.

4. Each of the supplied figure source data files and supporting gel images needs to be linked directly to a particular figure panel or panels. I have attached a zip file with these files, please change the file names according to the following example "Additional data for figure 2B.txt", and then send the zip file back to us. We already assigned the gel images to what seemed to be the appropriate figure panels; please double-check these assignments.

New files are be sent as attachments to reply email(s).

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

Sincerely,
Editor - Molecular Systems Biology
msb@embo.org

© European Molecular Biology Organization