



Published in final edited form as:

Cell Rep. 2021 December 21; 37(12): 110136. doi:10.1016/j.celrep.2021.110136.

## Gene network modeling via TopNet reveals functional dependencies between diverse tumor-critical mediator genes

Helene R. McMurray<sup>1,2</sup>, Aslihan Ambeskovic<sup>1</sup>, Laurel A. Newman<sup>1</sup>, Jordan Aldersley<sup>1</sup>, Vijaya Balakrishnan<sup>1</sup>, Bradley Smith<sup>1</sup>, Harry A. Stern<sup>3</sup>, Hartmut Land<sup>1,5,\*</sup>, Matthew N. McCall<sup>1,4,5,6,\*</sup>

<sup>1</sup>Department of Biomedical Genetics, University of Rochester Medical Center, 601 Elmwood Avenue, Rochester, NY 14642, USA

<sup>2</sup>Department of Pathology and Laboratory Medicine, University of Rochester Medical Center, 601 Elmwood Avenue, Rochester, NY 14642, USA

<sup>3</sup>Center for Integrated Research Computing, University of Rochester Medical Center, 601 Elmwood Avenue, Rochester, NY 14642, USA

<sup>4</sup>Department of Biostatistics and Computational Biology, University of Rochester Medical Center, 601 Elmwood Avenue, Rochester, NY 14642, USA

<sup>5</sup>Wilmot Cancer Institute, University of Rochester Medical Center, 601 Elmwood Avenue, Rochester, NY 14642, USA

<sup>6</sup>Lead contact

### SUMMARY

Malignant cell transformation and the underlying reprogramming of gene expression require the cooperation of multiple oncogenic mutations. This cooperation is reflected in the synergistic regulation of non-mutant downstream genes, so-called cooperation response genes (CRGs). CRGs affect diverse hallmark features of cancer cells and are not known to be functionally connected. However, they act as critical mediators of the cancer phenotype at an unexpectedly high frequency >50%, as indicated by genetic perturbations. Here, we demonstrate that CRGs function within a network of strong genetic interdependencies that are critical to the malignant state. Our network modeling methodology, TopNet, takes the approach of incorporating uncertainty in the underlying gene perturbation data and can identify non-linear gene interactions. In the dense space of gene connectivity, TopNet reveals a sparse topological gene network architecture, effectively

---

This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

\*Correspondence: land@urmc.rochester.edu (H.L.), mccallm@gmail.com (M.N.M.).

#### AUTHOR CONTRIBUTIONS

H.R.M., H.L., and M.N.M. conceived the study. H.R.M., A.A., L.A.N., J.A., V.B., and B.S. performed the gene perturbation experiments and analyzed the data from these. H.A.S. and M.N.M. developed the TopNet algorithm and software. M.N.M. and H.R.M. performed the computational and statistical analyses. H.R.M., A.A., H.L., and M.N.M. interpreted the results and wrote the manuscript.

#### SUPPLEMENTAL INFORMATION

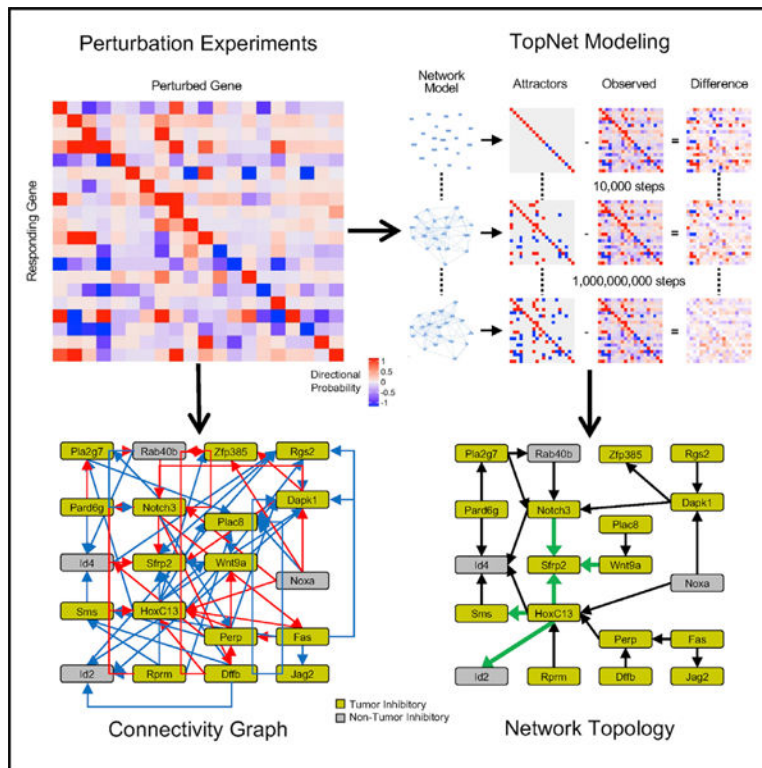
Supplemental information can be found online at <https://doi.org/10.1016/j.celrep.2021.110136>.

#### DECLARATION OF INTERESTS

The authors declare no competing interests.

pinpointing functionally relevant gene interactions. Thus, among diverse potential applications, TopNet has utility for identification of non-mutant targets for cancer intervention.

## Graphical abstract



## In brief

Malignant cell transformation requires the cooperation of multiple oncogenic mutations. Here, we demonstrate that non-mutated genes function within a network of strong genetic interdependencies that are critical to the malignant state. Our network modeling methodology, TopNet, reveals a sparse topological gene network architecture, effectively pinpointing functionally relevant gene interactions.

## INTRODUCTION

Oncogenic mutations are critical for both cancer initiation and maintenance, which has driven efforts to inhibit targetable oncoproteins as an effective strategy for cancer treatment. Targetable mutations, however, are found only in a small fraction of cancers. Thus, broader strategies for cancer intervention need to be developed, such as targeting non-mutated proteins or molecular circuitry essential to cancer cells. The complexity of cell regulation and the profound cellular reprogramming associated with malignant transformation, however, provide formidable barriers to elucidating disease-critical genetic circuitry. We propose that linking genetic perturbation experiments with statistical modeling can provide inroads to the discovery of functionally relevant gene regulatory network

(GRN) architecture and thus valuable identification and validation of genetic interactions in cancer gene networks, which ultimately may inform the targeting of next-generation cancer interventions.

GRN modeling seeks to infer dependencies between genes and thereby provide insight into the regulatory relationships that exist within a cell. Early methods to estimate GRNs focused primarily on data arising from targeted experimental perturbations (Ideker et al., 2000; Pe'er et al., 2001); however, these methods relied on simplifying assumptions to reduce the computational demands of these methods. As the size and scope of publicly available transcriptomic data rapidly increased due to the formation of large cancer consortia such as The Cancer Genome Atlas (TCGA), GRN modeling methodology transitioned to leveraging observational data, which rely on correlation structure in the data produced by natural biological covariation to infer regulatory relationships (Basso et al., 2005; Carro et al., 2010; Langfelder and Horvath, 2008; Margolin et al., 2006; Zhang and Horvath, 2005). While the application of these methods has provided insights into the regulatory systems crucial to malignancy, these modeling approaches are hindered by technical sources of covariation, such as batch effects or variable sample composition, either masking or mimicking true biological dependencies (Zhang et al., 2021). In addition, modeling of GRNs remains inherently difficult due to high-order non-linear interactions and unmeasured sources of biological variability that influence the GRN.

Methodological and computational advances have greatly aided our ability to model GRNs, which in turn has advanced our understanding of the processes involved in both normal cellular function and disease progression. While much effort has been dedicated to estimate GRNs from observational data, experimental perturbations allow direct measurement of the effect that a change in one gene has on other measured genes, which greatly facilitates the estimation of GRNs (Markowitz and Spang, 2003; Werhli et al., 2006; Zak et al., 2003). Recent advances in computational hardware and parallelization provide an opportunity to revisit the potential of perturbation-based GRN estimation without the need to impose biologically implausible restrictions on these models.

In this article, we propose a GRN modeling procedure, called TopNet, based on underlying gene perturbation data that addresses the core challenges of GRN estimation: preprocessing and analysis of noisy experimental data, quantification of uncertainty in gene expression changes following perturbation, exploration of a vast model space to find networks supported by the gene perturbation data, and identification of robust, biologically relevant network features. Our approach adopts the ternary network modeling formalism proposed by Almudevar et al. (2011). Specifically, we consider a ternary network modeling procedure that samples networks for which the network attractors agree with the experimental data. In this article, we propose improvements in both the search algorithm and the manner in which agreement between the attractors and observed data is quantified.

Previously, we were able to identify non-mutant mediator genes critical to the cancer phenotype downstream of oncogenic mutations (Ashton et al., 2012; McMurray et al., 2008). Our approach to find such mediators was based on the concept that the profound and multifaceted genetic reprogramming associated with malignant cell transformation requires

the cooperation of multiple oncogenic mutations (Hahn et al., 1999; Land et al., 1983), and that this transition is driven by synergistic modulation of downstream mediators (Lloyd et al., 1997; Sewing et al., 1997; Xia and Land, 2007). This class of functionally diverse and apparently unconnected genes was called cooperation response genes (CRGs). Functional analysis of CRGs revealed that >50% of the CRGs tested play an essential role in the cancer phenotype (McMurray et al., 2008) and that specific CRGs represented newly identified, cancer cell-specific vulnerabilities (Kinsey et al., 2014; Smith and Land, 2012; Smith et al., 2016). Given the functional diversity of CRGs and the robust and redundant nature of regulatory circuitry in mammalian cells, this high frequency of impact on the cancer phenotype by individual CRG perturbations is contrary to expectation, unless CRGs act in a concerted manner. Our work demonstrates that CRGs act within a strong network of unexpected genetic interdependencies that appears critical to the robustness of the malignant phenotype.

## RESULTS

### CRGs contribute to tumor formation capacity

Genetic perturbation of individual CRGs in cells transformed by p53<sup>175H</sup> and Ras<sup>V12</sup> (mp53/Ras cells), aimed to re-establish levels of gene expression found in non-transformed parental young adult mouse colon (YAMC) cells, results in a reduction of tumorigenicity in the majority of cases (McMurray et al., 2008). The extensive analyses presented here recapitulate our earlier results. We now show an entirely new and much larger set of 49 CRG perturbations (Figures 1, S1, and S2). Of these, 25 perturbations resulted in a significantly reduced tumor formation capacity of the target cells when implanted into immunocompromised mice. Together with our previous work, these data demonstrate that 43 of the 75 CRGs tested (57%, adjusted  $p < 0.1$ ) are critical to the cancer phenotype, including 2 perturbations that increase tumor size (Figure 1A).

CRGs regulate diverse hallmarks of the malignant phenotype. The distribution of associated cell processes among those genes with a demonstrated role in tumor regulation is highly similar to the distribution across all CRGs but distinct from known oncogenes and tumor suppressors (Figure 1B). The genes whose perturbation affects tumor formation vary in their localization from nuclear to intracellular to extracellular function and are associated with numerous regulatory pathways (Figure 1C). Notably, many of the CRGs that are implicated in the control of tumor formation capacity are associated with processes such as metabolism and cell adhesion, regulators of which are not detected among frequently mutated oncogenes and tumor suppressor genes. Thus, we consider the CRGs as critical mediators of oncogene cooperation that transmit and carry out the effects of those oncogenic mutations.

### Individual CRG perturbations affect expression of other CRGs

CRGs function in many diverse pathways and mechanisms (Figures 1B and 1C), and interrelationships between most of these genes have not been previously reported. In contrast, the high frequency by which CRG perturbations affect the cancer phenotype and the manner in which the CRGs themselves were identified, by virtue of their synergistic response to multiple oncogenic mutations, is consistent with the idea that CRGs function

within a regulatory gene network. We thus investigated the extent of mutual control of gene expression among 20 CRGs, which were selected to represent the spectrum of functional classes and tumor inhibitory effects observed in the full CRG set (indicated in Figures 1A and 1C). Using a set of mp53/Ras cell populations, each harboring 1 of 20 individual CRG perturbations, we compared the mRNA expression profiles of perturbed cells with corresponding controls to detect changes in the expression of the selected CRGs. These perturbations varied in their effects on tumor growth (Figure 1A) and were chosen from genes up- or downregulated in transformed mp53/Ras cells as compared to the parental YAMC cells (Figure 2A). Expression profiles for all 20 CRGs were measured for multiple independent replicates of each CRG perturbation, revealing widespread changes in gene expression in the CRG cohort upon perturbation of individual CRGs (Figure 2B).

The high degree of connectivity between the selected CRGs is shown in Figures 2C and S3A. Key parameters that describe the interconnectedness of genes in such a graph are the out-degree (i.e., the number of children of a given gene) and the in-degree (i.e., the number of parents of a given gene). Of the 20 perturbations, 8 produced changes in the expression of 5 of the other CRGs, with 1 perturbation, restoration of HoxC13 expression, having the highest out-degree (i.e., affecting the expression of 9 other CRGs) (Figures S3A and S3B). Only two perturbations had an out-degree of zero, indicating that they did not produce a change in any other CRG in the network. Moreover, there was substantial variability in the in-degree among these CRGs, ranging from two CRGs not responding to any other perturbations, to the one CRG, *Sfrp2*, with the highest in-degree, responding to the perturbation of 8 other CRGs (Figures S3A and S3C). These results highlight the remarkably high connectedness of these 20 CRGs. Given the numerous interactions between this subset of CRGs, we needed to move beyond simple connectivity analyses to prioritize interactions for biological testing. Thus, we used statistical modeling to understand the flow of information in this gene regulatory network and estimate the associated network architecture.

### TopNet reveals rules governing attractor states

To identify gene interactions relevant to the malignant phenotype, we developed TopNet, a ternary network modeling approach based on perturbations capable of up- or downregulation of gene expression that accounts for uncertainty in the underlying gene perturbation data. This appeared suitable, as CRGs comprise both up- and downregulated genes, including genes with very low expression levels (McMurray et al., 2008).

The estimation of a GRN typically focuses on inferring individual interactions (edges) between genes and/or proteins (nodes). The collective behavior of these interactions is then studied as an emergent property of the low-level network architecture. This is a natural strategy when inferring GRNs based on the measurement of pairwise interactions—for example, transcription factor binding (either via ChIP-seq or promoter sequence motifs) or protein-protein interaction data. However, gene perturbation experiments remain the most reliable approach to investigate GRNs and to test cellular responses to interventions (Markowitz and Spang, 2003; Werhli et al., 2006; Zak et al., 2003). The gene expression changes in response to perturbation provide information not on the direct

interactions between genes but rather on the altered steady-state expression produced by the perturbation. In other words, the estimation procedure can be framed as a search through the vast space of network models for those that produce the observed data (Almudevar et al., 2011).

To assess potential interdependence between the 20 CRGs selected for network estimation, we applied the TopNet algorithm, which uses a ternary network model that accounts for the dynamic, cyclic, and non-linear nature of GRNs and facilitates the evaluation of uncertainty (Almudevar et al., 2011). Specifically, TopNet assumes that changes in gene expression can be modeled by one of three states—downregulation (−), baseline expression (0), or upregulation (+1)—and that the network is governed by deterministic transition functions that encode the complex regulatory relationships between genes (examples are shown in Figures 3A–3D). Discretization captures the majority of information and inherently provides robustness (McCall et al., 2011; Parmigiani et al., 2002; Scharpf et al., 2003; Zilliox and Irizarry, 2007). Additional details of the TopNet modeling algorithm are described in Method details, and Methods S2 and S3.

Following any given perturbation, the network will eventually reach a steady state, called an attractor. This attractor may be a new state to which the cell is driven by the action of the gene perturbation, or it may represent a return to the baseline state—in this case that of a transformed mp53/Ras cell. Recall that the experimental data represent a measurement of the steady-state expression of the cells; therefore, the ability of a network model to represent experimental data can be assessed by comparing these gene expression measurements with the attractors produced by the network model (Almudevar et al., 2011). Specifically, starting from a null model, TopNet randomly proposes small changes in the network structure and scores the proposed new network based on the similarity between its attractors and the observed steady-state data (Figure 4; Method details, and Methods S1, S2, and S3). This iterative search algorithm can proceed quite rapidly because TopNet can easily compute the attractor for each perturbation and assess the similarity of the resulting attractors with the corresponding observed steady-state data. Ultimately, TopNet identifies a network model whose attractors match the observed data well.

### **Uncertainty quantification allows the identification of high-confidence features of the network topology**

There are often many network models that fit the observed data equally well, and the model space is huge; for the network model presented here, there are  $\sim 2.48 \times 10^{763}$  potential networks. To more efficiently explore this vast model space and avoid becoming trapped in local optima, we implemented a replica exchange Monte Carlo algorithm (Swendsen and Wang, 1986) that allowed us to parallelize the search over the nodes of a high-performance computing cluster. Unlike previous approaches, here, we have modeled and addressed non-random missing data (McCall et al., 2014), batch effects, and uncertainty in estimates of differential expression (Methods S1 and S2).

Due to the vast number of potential networks, it is exceedingly unlikely that there exists a single “best” network; therefore, as previously proposed by Almudevar et al. (2011), TopNet repeats the estimation procedure many times to produce a sample of networks from



the space of models that explain the observed data well. This collection of network fits can then be used to quantify the degree of support for a given aspect of the network by calculating the proportion of networks in which a given feature or features are present. For example, based on the 4 network models shown in Figure 3, we could state that HoxC13 is likely a direct parent of Sms because that relationship is present in all 4 models, whereas we have less certainty about the relationship between HoxC13 and Id2 because it is only a direct parent in 3 of the 4 models, remembering that the 4 models shown are only a subsample of all that were computed.

Building upon previous work (Almudevar et al., 2011), TopNet implements several advances in modeling and computation (see Method details). These methodological improvements, including quantification of uncertainty, produced more reliable estimates of the network topology than previous methods (Figure S4 and Methods S2). In addition, we show that the selected in-degree threshold of 4 is flexible enough to produce network models that explain the observed data well without being so flexible that even random data can be modeled well (Figure S5 and Methods S2). Finally, we show general agreement in network topology for sufficiently complex models (Figure S6 and Methods S2). These improvements to data analysis and network modeling were crucial to identify high-confidence features of the network topology.

### **Modeling reveals robust topological features of the CRG regulatory network**

The result of the TopNet network modeling procedure is a sample of network models, all of which explain the observed data well. Based on this sample of network models, we are able to form inferential statements about specific network features by averaging across the network models. For example, in the case of topological relationships, we calculate the proportion of networks in which a given gene is a parent of another gene. For the CRG network, we report results based on 100 independent network fits. In Figure 5, we show edges present in at least 80% of the network models, which represent regulatory relationships that are strongly supported by the observed data.

The network model provides both a sparse topology of gene interactions (Figure 5) and direction of gene responses, derived from the underlying transition functions (see examples in Figure 3). Based on this information, we are able to prioritize genetic perturbation experiments testing particular gene interactions and their role in maintaining or generating a given biological state (i.e., the malignant state). Here, we have focused on investigating the impact of multiple pairwise gene interactions relating to two central players common to all of the network models, HoxC13, the parent node with the most children, and Sfrp2, the gene affected by the largest proportion of perturbations (Figures 2C and S3).

### **Interdependency of functionally diverse mediator genes**

To test whether the CRG parent-child relationships suggested by the network model reflect functional interdependencies between parent-child pairs, we chose an example in which perturbation of the upstream gene has a reproducible biological effect—in this case, inhibition of tumor growth *in vivo*. Notably, genetic perturbation (i.e., re-expression of HoxC13, a homeobox transcription factor) (Godwin and Capecchi, 1998; McMurray et al., 2008),

produced some of the strongest antitumor effects observed in our experiments. Next, we carried out double perturbations of parent-child pairs, aimed to limit the impact of perturbing parent gene expression. For double perturbations, we selected the parent-child relationships between HoxC13 and its high-confidence children: Sms, the gene encoding spermine synthase (Pegg and Michael, 2010); Id2, a gene encoding a member of the helix-loop-helix only and inhibitor of DNA-binding protein family (Lasorella et al., 2001); and Sfrp2, encoding a member of the family of secreted Frizzled-related proteins, antagonists of Wnt signaling (Rattner et al., 1997) (Figure 6A). Note that the relationship between HoxC13 and Id4, while specified by the network topology, is not observed in the connectivity graph, and thus an appropriate genetic perturbation to test this relationship remains unclear.

Our experiments testing relationships between HoxC13 and Sms, Id2, or Sfrp2 demonstrate that changes to the expression of these child nodes are required for the biological effects of the parent perturbation (Figures 6B–6D and S7). All of these relationships represent previously unknown genetic interdependencies, or functional epistatic relationships, critical to the cancer phenotype. Specifically, we find that the ability of HoxC13 to suppress the malignant state is curtailed when either low Sfrp2 levels or high Sms or Id2 levels are enforced. While double perturbation of HoxC13 with either Sms or Id2 appears to completely rescue tumor formation, combining HoxC13 re-expression with Sfrp2 knockdown shows only a partial rescue. This may relate to additional complexity in the relationship between HoxC13 and Sfrp2, with other molecules playing a role in the phenotypic output, or it may relate to technical limitations of the experimental protocol. Our results, however, are sufficiently robust that they indicate biologically relevant gene interactions.

The GRN presented suggests specific interactions that may help to maintain the robustness of the network and ultimately the malignant state of the cells. However, parent-child relationships beyond those identified by modeling the GRN can be seen by direct examination of the impact of a perturbation on the other genes in consideration (Figures 2 and 6A). Specific to HoxC13 as a parent, gene expression profiling reveals that the expression of Plac8, encoding a lysosomal protein critical for autophagosome-lysosome fusion in cancer cells (Galaviz-Hernandez et al., 2003; Kinsey et al., 2014), and Rgs2, which encodes a family member of regulators of G protein-coupled receptors (Kannangai et al., 2007), are each decreased by HoxC13 perturbation. However, these interactions are not necessary for the GRN to explain the steady-state observations of expression of all of the genes in the network and thus are not frequently observed in the topological mapping of the GRN. To test the relative importance of interactions identified by network modeling versus those observed by gene expression profiling, we generated double perturbations re-expressing HoxC13 together with either Plac8 or Rgs2. Remarkably, neither of these combinations revealed the dependence of the antitumor effects of HoxC13 on Plac8 or Rgs2 (Figures 6E and S7), although enforcing both HoxC13 and Plac8 expression had an enhanced antitumor effect, an interesting observation worthy of study in a different context. Testing of the biological dependence of HoxC13 on potential interactions suggested by the network model and those observed exclusively in the connectivity graph shows that the GRN prioritizes biologically relevant genetic interdependencies (Figure 5) among the densely populated space of relationships found in the connectivity graph (Figure 2B).



Furthermore, we examined whether relationships between *Sfrp2* and its parents, *HoxC13*, *Notch3*, and *Wnt9a* (Figures 5 and 6A) are critical to the malignant phenotype. As described above, the re-expression of *HoxC13* does not significantly inhibit tumor growth when low *Sfrp2* levels are maintained. Similarly, the tumor inhibitory effect of *Notch3* expression is suppressed when *Sfrp2* expression is maintained at low levels (Figures 6F and S7). The re-expression of *Wnt9a* has a small but statistically significant effect on tumor size (Figures 1A and 6G). In contrast to *HoxC13* and *Notch3* re-expression, increased *Wnt9a* expression decreases *Sfrp2* mRNA levels (Figure 2A). When *Sfrp2* and *Wnt9a* expression were both increased, we observed a modest but significant reduction in tumor growth compared to controls ( $p = 0.029$ , unadjusted Wilcoxon signed-rank test) but increased tumor growth compared to *Sfrp2* expression alone (Figures 6G and S7). This observation is consistent with the idea that *Sfrp2* and *Wnt* antagonize each other both at the levels of gene expression and *Wnt* signaling activity (Rattner et al., 1997).

Here, we demonstrate that CRGs function within a gene regulatory network that contributes to the malignant phenotype. Our results reveal that numerous tumor regulatory interactions are detected via attractor-based ternary GRN modeling through the TopNet algorithm. Notably, high-confidence topological features of the network model are frequently found to be biologically relevant, and thus the network model effectively pinpoints previously unknown gene interdependencies critical to the cancer phenotype.

## DISCUSSION

Here, we show that apparently unrelated mediators of the cancer phenotype are linked by a strong network of genetic interactions and thus maintain robustness of the malignant state. Critical to this discovery, we developed and applied TopNet, an approach to GRN estimation capable of identifying a biologically relevant topological model of highly complex genetic interdependencies. In contrast to the connectivity graph, TopNet can accurately model cellular responses to perturbations, produce testable hypotheses regarding non-linear multi-gene interactions, and identify control points in the network architecture. This involves searching through a massive number of potential network models to identify those instances consistent with the observed data. By combining the information from many network models, TopNet prioritizes aspects of the GRN that are most strongly supported by the experimental data.

TopNet is capable of pinpointing key architectural features of cancer cells. The application of TopNet to genetic perturbation experiments involving 20 CRGs has identified high-confidence multi-gene interactions involved in the maintenance of the cancer phenotype. Furthermore, experimentation confirmed all of the tested interactions identified by TopNet as previously unknown genetic interdependencies. In fact, all of the relationships between *HoxC13* and its children *Sms*, *Sfrp2*, or *Id2*, reported here, were experimentally confirmed to be contributing to the tumor-inhibitory activity of *HoxC13*, which is downregulated in transformed cells as compared to non-transformed parental controls (McMurray et al., 2008). Our data suggest that *HoxC13* inhibits the pathways downstream of *Sms* and *Sfrp2* (i.e., polyamine metabolism and *Wnt* signaling) and *Id2* function, all of which is consistent with our observation of reduced tumor growth capacity of transformed cells in which

HoxC13 expression has been elevated to resemble levels observed in non-transformed parental cells. Similarly, we demonstrated functional relevance for all of the interactions between Sfrp2 and its high-confidence parents Notch3 and Wnt9A, indicating that the tumor-inhibitory activity of Notch3 is dependent on Sfrp2 and that Wnt9A can downregulate Sfrp2 expression and thus neutralize Sfrp2-mediated tumor inhibition, consistent with the activation of Wnt signaling. We expect that further experimentation, especially combined perturbations of >2 genes, would likely reveal additional complexity. The feasibility of such studies, however, is limited by current technologies.

In addition to identifying high-confidence dependencies via examination of the network topology, TopNet provides the corresponding transition functions that determine the nature of dependencies between parent and child nodes (e.g., Figure 3). However, one should not assume that either the summarized network topology or the transition functions necessarily represent direct mechanistic relationships in the cell. In fact, the functional diversity of the 20 CRGs examined in this work suggests that most of the identified interactions and dependencies act through unmeasured intermediaries. Nevertheless, armed with information on critical interactions in the cell, one can then dedicate effort to understanding such interactions at a mechanistic level.

A major strength of our work is the use of data from gene perturbation experiments as the input for TopNet modeling. While a number of network modeling approaches have been proposed to identify and understand epistatic interactions in cancer cells, many of these modeling approaches are hindered by being observational in nature, as they rely on correlation structure in the data produced by natural biological covariation (van de Haar et al., 2019). In contrast, network modeling based on perturbation experiments allows prioritization of cause-effect relationships before experimental validation. This has led to increased interest in perturbation experiments as a means of modeling biological systems, such as the use of Perturb-seq (Adamson et al., 2016; Dixit et al., 2016; Jaitin et al., 2016; Schraivogel et al., 2020) to measure the effect of perturbations at single-cell resolution.

While high-throughput perturbation approaches such as Perturb-seq are currently limited by their reliance on gene extinction as the means for gene perturbation, as well as the reduced sensitivity of single-cell RNA sequencing technology, we are intrigued by the possibility of adapting our GRN estimation methodology to larger and potentially more complex networks. The extension of TopNet to such data would allow one to move beyond associative analyses and identify non-linear gene interactions underlying cancer cell robustness. Several computational and statistical challenges will need to be addressed in this context, such as the increased feature space, decreased sensitivity, and unidirectional perturbations. In the near term, targeted Perturb-seq (Schraivogel et al., 2020), in which some of these challenges are lessened, is ideally suited for initial methods development and testing. Our general modeling approach has the potential to extract rich biological interactions from these data and open new avenues of biological inquiry.

The complexity and strength of the CRG network revealed by TopNet are both remarkable and unexpected. CRGs comprise a diverse set of genes functioning in a multitude of diverging pathways involved in regulating cell signaling, metabolism, gene expression,

adhesion, and survival. In addition, biologically relevant genetic interactions between any of the 20 CRGs investigated here, to our knowledge, have not been reported. Our observations thus suggest a fundamental role for genetic networks in supporting the robustness of the malignant state and reveal network stability per se as a target for cancer intervention. Combination perturbations that disrupt network robustness may be needed for successful intervention. One could imagine that such an approach would be advantageous in the context of emerging treatment resistance. In addition, the examination of gene regulatory network architecture also has the potential to uncover an unexpected wealth of cancer cell vulnerabilities controlled by non-mutated genes and ultimately identify druggable targets to guide novel intervention strategies. For example, targeting a metabolic enzyme such as Sms could serve as a convenient alternative to intervention at the level of HoxC13, as inhibition of Sms expression is similarly effective in tumor inhibition as HoxC13 reactivation. In addition, molecular mechanisms underlying the Sms-HoxC13 interaction would be worth exploring further. TopNet, as shown here, thus can effectively be used for the identification of suitable targets accessible to pharmacological intervention in a largely non-druggable GRN environment.

## STAR★METHODS

### RESOURCE AVAILABILITY

**Lead contact**—Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Matthew McCall (matthew\_mccall@urmc.rochester.edu).

**Materials availability**—Plasmids generated in this study are available upon request.

**Primer sequences and TaqMan probe sets are tabulated in Tables S1–S3.**—There are restrictions to the availability of YAMC cells due to materials transfer agreement.

### Data and code availability

- All qPCR data are available in this paper's supplemental information. Tumor volume data reported in this paper will be shared by the lead contact upon request.
- All original code to reproduce all statistical analyses is available in the crgnet R package included in this paper's supplemental information as Methods S1 and the ternarynet R/Bioconductor package, available at: <https://bioconductor.org/packages/release/bioc/html/ternarynet.html>.
- The pseudocode for the network modeling algorithm is supplied in Methods S3, while a reproducible workflow of the analyses in this paper is included as the vignette of the crgnet package (Methods S1).
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

## EXPERIMENTAL MODEL AND SUBJECT DETAILS

**Cells**—Low-passage polyclonal young adult mouse colon (YAMC) cells were originally derived from males and females of the Immorto-mouse (H-2K<sup>b</sup>/tsA58 transgenic mouse) and gifted to the Land laboratory by a gift from R. Whitehead and A.W. Burgess (D'Abaco et al., 1996; Whitehead et al., 1993). YAMC cells express temperature sensitive simian virus 40 large T (tsA58) under the control of an interferon  $\gamma$  inducible promoter were maintained at the permissive temperature (33°C) for large T in the presence of interferon  $\gamma$  to support conditional immortalization *in vitro*. This permits expansion of the cells in tissue culture. In contrast, exposure of YAMC cells to the non-permissive temperature for large T (39°C) in the absence of interferon  $\gamma$  leads to growth arrest followed by cell death, indicating the absence of spontaneous immortalizing mutations in the cell population. The cells were cultured on Collagen IV-coated dishes (1 $\mu$ g/cm<sup>2</sup> for 1.5 hr at room temp; Sigma) in RPMI 1640 medium (Invitrogen) containing 10% (v/v) fetal bovine serum (FBS) (Hyclone), 1  $\times$  ITS-A (Invitrogen), 2.5  $\mu$ g/ml gentamycin (Invitrogen), and 5 U/ml interferon  $\gamma$  (R&D Systems). All experiments testing the effects of Ras<sup>V12</sup> and p53<sup>175H</sup> were carried out at the non-permissive temperature for large T function (39°C) and in the absence of interferon  $\gamma$ .

**Animals for allograft tumor assays**—Tumor formation was assessed by sub-cutaneous injection of perturbed or control cells into nulliparous female adult CD-1 nude mice (Crl:CD-1-Foxn1<sup>nu</sup>, Charles River Laboratories, purchased at 6 – 8 weeks of age). Animals were housed 3 – 5 per cage in micro-isolator conditions, necessitated by their immunocompromised status. Cells were implanted at a dosage of 5 $\times$ 10<sup>5</sup> per injection in RPMI 1640 media with no additives. For each gene perturbation three to six biological replicates were performed; 4 – 8 injections were performed for each replicate of perturbed cells and vector controls. Two cell implantations were performed on each mouse in the left and right flanks; animals received exclusively perturbed cells or vector control cell populations. Tumor size was measured by caliper at 2, 3 and 4 weeks post-injection. Tumor volume was calculated by the formula volume = (4/3) $\pi$ r<sup>3</sup>, using the average of two radius measurements. Tumor reduction was calculated based on the difference of average tumor volume following each gene perturbation as compared to the vector control tumors from the same batch of perturbations. Statistical significance of difference in average tumor size was calculated by the Wilcoxon signed-rank test with Benjamini-Hochberg multiple testing correction (Hochberg and Benjamini, 1990).

## METHOD DETAILS

**Cells**—Young adult mouse colon (YAMC) cells and derivation of transformed cells with multiple oncogenic lesions were derived and are used as described elsewhere (McMurray et al., 2008; Whitehead et al., 1993). Briefly, polyclonal cell populations harboring mutant forms of p53 (p53<sup>175H</sup>) and Ha-Ras (Ras<sup>V12</sup>) (abbreviated as mp53/Ras) were derived by retroviral infection of low-passage polyclonal YAMC cells. YAMC cells were cultured on Collagen IV-coated dishes (1  $\mu$ g/cm<sup>2</sup> for 1.5 hr at room temp; Sigma) in RPMI 1640 medium (Invitrogen) containing 10% (v/v) fetal bovine serum (FBS) (Hyclone), 1  $\times$  ITS-A (Invitrogen), 2.5  $\mu$ g/ml gentamycin (Invitrogen), and 5 U/ml interferon  $\gamma$  (R&D Systems). All experiments testing the effects of genetic perturbations were carried out at the non-permissive temperature for large T function (39°C) and in the absence of interferon  $\gamma$ .

**Genetic perturbation of gene expression**—cDNAs expressed via pBabe retroviral vectors or pLenti6 lentiviral vectors and shRNA delivered via pSuper-retro retroviral vectors or pLKO lentiviral vectors were used to generate gene perturbations. These were tested by comparison of RNA expression levels in empty vector-infected cells and cells subjected to gene perturbation via SYBR Green qPCR with gene-specific primers.

cDNA clones were obtained from the IMAGE consortium collection, distributed by Open Biosystems, or PCR-cloned from murine cDNA using sequence-specific primers except for murine Jagged2 (Jag2) and Notch3-intracellular domain (Notch3-ICD) (gifts of Dr. L. Milner). All cDNAs were sequence-verified prior to use and were cloned into the retroviral vector pBabe-puro (Morgenstern and Land, 1990).

shRNA molecules were designed using an algorithm (Yuan et al., 2004), available at <http://sirna.wi.mit.edu/home.php>, for cloning into the pSuper-retro vector (Oligoengine) or purchased as clones in the pLKO vector. Target sequences for pSuper-retro cloning were synthesized as forward and reverse oligonucleotides (IDT), which were annealed and cloned into the vector. For each upregulated gene, we identified two or three independent shRNA target sequences yielding at least 50% reduction in gene expression with the goal to guard against off-target effects (Figure S2). For this purpose between four and six shRNA targets for each gene were tested. In the case of SerpinB2, only one shRNA target sequence yielded appropriate levels of knock-down, reducing mRNA expression to levels comparable to those in YAMC cells.

Retroviruses were produced following transient transfection of  $\Phi$ NX-eco cells, while lentiviruses were produced following transient transfection of 293T cells. Infections were carried out in media with 8  $\mu$ g/mL polybrene. Selection with 5  $\mu$ g/mL puromycin, and where applicable, 200  $\mu$ g/mL hygromycin B, was used to generate polyclonal populations of cells stably expressing the indicated cDNAs and/or shRNAs. To test reproducibility of the effects of CRG gene perturbations on tumor formation 2–4 independent replicates of such cell populations were derived (Figure S2).

For combined perturbations, cDNA or shRNA for one gene in the pair was sub-cloned into the appropriate pBabe-hygro or pSuper-retro-hygro retroviral vector, allowing for consecutive, independent selection for each gene perturbation introduced.

**Quantitation of gene perturbation**—The efficiency of gene perturbations was tested by comparison of RNA expression levels in empty vector-infected mp53/Ras cells and cells subjected to gene perturbation. Re-expression or knock-down was also compared with the respective levels of RNA expression in YAMC control cells. For collection of RNA, mp53/Ras cells were grown at the 39°C for 2 days, followed by serum withdrawal for 24 hr. Total RNA was extracted from cells following the standard RNeasy Mini Kit protocol for animal cells, with on-column DNase digestion (QIAGEN).

SYBR Green-based quantitative PCR was run using cDNA produced as described below for TLDA, with 1x Bio-Rad iQ SYBR Green master mix, 0.2  $\mu$ M forward and reverse primer mix, with gene-specific qPCR primers for each gene tested. Primers were identified using

the Primer Bank database<sup>15</sup>, available at <https://pga.mgh.harvard.edu/primerbank/index.html> or designed using the IDT PrimerQuest tool (<https://www.idtdna.com/Scitools/Applications/Primerquest/>). Differential gene expression was calculated by the Ct method, described below, using RhoA as an endogenous reference gene. Reactions were run on the iCycler (Bio-Rad), as follows: 5 min at 95°C, 45 cycles of 95°C for 30 s, 58 to 61°C for 30 s, 68 to 72°C for 45 s to amplify products, followed by 40 cycles of 94°C with 1°C step-down for 30 s to produce melt curves.

As expected, the magnitude of perturbation varies between cDNAs and replicates, and falls into the following groups. For tumor-inhibitory CRGs, all replicates express cDNAs at levels below, at or moderately above YAMC mRNA expression levels with the exception of Pvr14, for which we cannot exclude the possibility that its tumor inhibitory effects are due to overexpression of the cDNA. For non-tumor-inhibitory CRGs, cDNA expression levels were found at or above the levels of the corresponding YAMC mRNAs (Figure S2).

**Allograft assays**—Tumor formation was assessed by sub-cutaneous injection of cells into CD-1 nude mice (CrI: CD-1-Foxn1<sup>tm</sup>, Charles River Laboratories). Murine mp53/Ras cells were grown at 39°C for 2 days prior to injection and implanted via sub-cutaneous injection of  $5 \times 10^5$  mp53/Ras in RPMI 1640 with no additives. For each gene perturbation three to six biological replicates were performed; 4 – 8 injections were performed for each replicate of perturbed cells and vector controls. Tumor size was measured by caliper at 2, 3 and 4 weeks post-injection. Tumor reduction was calculated based on the difference of average tumor volume following each gene perturbation as compared to the vector control tumors from the same batch of perturbations. Statistical significance of difference in average tumor size was calculated by the Wilcoxon signed-rank test with Benjamini-Hochberg multiple testing correction (Hochberg and Benjamini, 1990).

## TLDA QPCR

Expression values for each of the CRGs were derived from TaqMan Low-Density QPCR Array (TLDA) data. The TaqMan Low-Density Array (Applied Biosystems) consists of TaqMan qPCR reactions representing the cooperation response genes and control genes (18S rRNA, GAPDH) in a microfluidic card. TLDA were used to measure gene expression patterns following genetic perturbation of CRGs. To generate cDNA for qPCR analysis, quadruplicate samples of mRNA from mp53/Ras cells harboring CRG perturbations were isolated and 10 µg/sample were mixed with 1x SuperScript II reverse transcriptase buffer, 10 mM DTT, 400 µM dNTP mixture, 0.3 ng random hexamer primer, 2 µL RNaseOUT RNase inhibitor and 2 µL of SuperScript II reverse transcriptase in a 100 µL reaction (all components from Invitrogen). RT reactions were carried out by denaturing RNA at 70°C for 10 minutes, plunging RNA on to ice, adding other components, incubating at 42°C for 1 hour and heat inactivating the RT enzyme by a final incubation at 70°C for 10 minutes.

For each sample, 82 µL of cDNA was combined with 328 µL of nuclease free water (Invitrogen) and an equal volume of TaqMan Universal PCR Master Mix No AmpErase UNG (Applied Biosystems). The mixture was loaded into each of 8 ports on the card at 100 µL per port. Each reaction contained forward and reverse primer at a final concentration



of 900 nM and a TaqMan MGB probe (6-FAM) at 250 nM final concentration. The cards were sealed with a TaqMan Low-Density Array Sealer (Applied Biosystems) to prevent cross-contamination. The real-time RT-PCR amplifications were run on an ABI Prism 7900HT Sequence Detection System (Applied Biosystems) with a TaqMan Low Density Array Upgrade. Thermal cycling conditions were as follows: 2 min at 50°C, 10 min at 94.5°C, 40 cycles of 97°C for 30 s, and annealing and extension at 59.7°C for 1 minute. Each individual replicate cDNA sample was processed on a separate card.

## QUANTIFICATION AND STATISTICAL ANALYSIS

**TopNet:** Our approach to topological network analysis consists of the following specific steps:

1. Non-detect imputation through application of an ECM algorithm to non-random missing data (McCall et al., 2014)
2. Calculation of z-scores for each perturbation based on Ct values to standardize expression fold changes.
3. Calculation of the probability of up- or downregulation in response to each perturbation by fitting a uniform/normal/uniform mixture model to the z-scores calculated in step 2.
4. Identification of ternary network models (Almudevar et al., 2011) which minimize the  $L_1$  distance between the network attractors and the estimated probabilities from step 3. For the network model presented in this manuscript there are approximately  $2.48 \times 10^{763}$  potential networks, so we use a specific implementation of replica exchange Monte Carlo (Swendsen and Wang, 1986) to search the model space in parallel across the nodes of a high-performance computing cluster.
5. Summarization of network topology by reporting parent-child relationships that are present in the vast majority, here at least 80 out of 100, of independent network fits.
6. Generation of testable hypotheses regarding interdependencies among genes in the network via examination of the network topology and the biological effects of perturbation of the genes in the network, in this manuscript inhibition of tumor growth *in vivo*.

Several of these steps are described in more detail in the following subsections.

**Preprocessing of gene expression data**—Gene expression values were derived using SDS 2.0 and 3.0 software packages (Applied Biosystems). As previously reported in other datasets, we observed a strong dependency between the proportion of non-detects (those reactions failing to produce fluorescence values above a certain threshold) and the average observed expression value. Non-detects were treated as non-random missing data and imputed using the R/Bioconductor package non-detects (McCall et al., 2014). The R/Bioconductor HTqPCR package (Dvinge and Bertone, 2009) was used to normalize the data to *Becn1* expression, which was shown to have relatively low variability across replicate

control samples, and accounts for much of the technical variability between samples (detailed in Methods S1).

Differential gene expression was calculated by the  $\Delta Ct$  method. Briefly, using threshold cycle (Ct) for each gene, change in gene expression was calculated for each sample comparison by the formulae:

$$\Delta Ct_{(\text{test sample})} = Ct_{(\text{target gene, test sample})} - Ct_{(\text{reference gene, test sample})}$$

$$\Delta Ct_{(\text{control sample})} = Ct_{(\text{target gene, control sample})} - Ct_{(\text{reference gene, control sample})}$$

$$\Delta \Delta Ct = \Delta Ct_{(\text{test})} - \Delta Ct_{(\text{calibrator})}$$

In order to incorporate uncertainty into subsequent analyses, we estimate the probability that a gene is up- / downregulated in response to each perturbation by fitting a uniform / normal / uniform mixture model to approximate z-scores for each perturbation. This approach is similar to the probability of expression (POE) algorithm (Parmigiani et al., 2002).

**Gene regulatory network reconstruction**—The methodology used to estimate the GRN uses the theoretical framework developed in Almudevar et al. (2011). One aspect of TopNet’s approach to GRN modeling is a method of scoring the networks based on a measure of uncertainty in the differential expression estimates. Specifically, the network models are scored by summing the absolute value of the differences between probabilities of up- or downregulation and the attractors and subtracting the best theoretically possible score. This allows the network model to give more weight to data points with higher certainty (probabilities closer to 0 or 1). As an added benefit, the ability to produce non-integer network scores eased transitions between network models and significantly decreased computational time. Another aspect of TopNet is an improved search algorithm that uses replica exchange Monte Carlo (Swendsen and Wang, 1986) to parallelize the search across many nodes of a high-performance computing cluster. This allowed us to avoid becoming trapped in local optima and to fit larger more complex networks in substantially less time. Results presented in this manuscript are based on network fits each run for 1,000,000,000 cycles in parallel across 20 processors with temperatures ranging from 0.001 to 1. We have also examined the transition functions, attractors, and trajectories all stored in the fits object, available within Methods S1.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## ACKNOWLEDGMENTS

We would like to thank the Wilmot Cancer Institute Genomics Shared Resource and Biostatistics and Bioinformatics Shared Resource and the Center for Integrated Research Computing at the University of Rochester

for technical support. The project described in this publication was supported by the National Human Genome Research Institute of the National Institutes of Health under award no. HG006853 (to M.N.M.) and the National Cancer Institute of the National Institutes of Health under award nos. CA138249 and CA197562 (to H.L.). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

## REFERENCES

- Adamson B, Norman TM, Jost M, Cho MY, Nunez JK, Chen Y, Villalta JE, Gilbert LA, Horlbeck MA, Hein MY, et al. (2016). A Multiplexed Single-Cell CRISPR Screening Platform Enables Systematic Dissection of the Unfolded Protein Response. *Cell* 167, 1867–1882.e21. [PubMed: 27984733]
- Almudevar A, McCall MN, McMurray H, and Land H (2011). Fitting Boolean networks from steady state perturbation data. *Stat. Appl. Genet. Mol. Biol* 10, 47.
- Ashton JM, Balys M, Neering SJ, Hassane DC, Cowley G, Root DE, Miller PG, Ebert BLM, McMurray HR, Land H, and Jordan CT (2012). Gene sets identified with oncogene cooperativity analysis regulate in vivo growth and survival of leukemia stem cells. *Cell Stem Cell* 11, 359–372. [PubMed: 22863534]
- Basso K, Margolin AA, Stolovitzky G, Klein U, Dalla-Favera R, and Califano A (2005). Reverse engineering of regulatory networks in human B cells. *Nat. Genet* 37, 382–390. [PubMed: 15778709]
- Carro MS, Lim WK, Alvarez MJ, Bollo RJ, Zhao X, Snyder EY, Sulman EP, Anne SL, Doetsch F, Colman H, et al. (2010). The transcriptional network for mesenchymal transformation of brain tumours. *Nature* 463, 318–325. [PubMed: 20032975]
- D’Abaco GM, Whitehead RH, and Burgess AW (1996). Synergy between Apc min and an activated ras mutation is sufficient to induce colon carcinomas. *Mol. Cell. Biol* 16, 884–891. [PubMed: 8622690]
- Dixit A, Parnas O, Li B, Chen J, Fulco CP, Jerby-Aron L, Marjanovic ND, Dionne D, Burks T, Raychowdhury R, et al. (2016). Perturb-Seq: Dissecting Molecular Circuits with Scalable Single-Cell RNA Profiling of Pooled Genetic Screens. *Cell* 167, 1853–1866.e17. [PubMed: 27984732]
- Dvinge H, and Bertone P (2009). HTqPCR: high-throughput analysis and visualization of quantitative real-time PCR data in R. *Bioinformatics* 25, 3325–3326. [PubMed: 19808880]
- Galaviz-Hernandez C, Stagg C, de Ridder G, Tanaka TS, Ko MS, Schlessinger D, and Nagaraja R (2003). Plac8 and Plac9, novel placental-enriched genes identified through microarray analysis. *Gene* 309, 81–89. [PubMed: 12758124]
- Godwin AR, and Capecchi MR (1998). Hoxc13 mutant mice lack external hair. *Genes Dev* 12, 11–20. [PubMed: 9420327]
- Hahn WC, Counter CM, Lundberg AS, Beijersbergen RL, Brooks MW, and Weinberg RA (1999). Creation of human tumour cells with defined genetic elements. *Nature* 400, 464–468. [PubMed: 10440377]
- Hochberg Y, and Benjamini Y (1990). More powerful procedures for multiple significance testing. *Stat. Med* 9, 811–818. [PubMed: 2218183]
- Ideker TE, Thorsson V, and Karp RM (2000). Discovery of regulatory interactions through perturbation: inference and experimental design. *Pac. Symp. Biocomput* 2000, 305–316.
- Jaitin DA, Weiner A, Yofe I, Lara-Astiaso D, Keren-Shaul H, David E, Salame TM, Tanay A, van Oudenaarden A, and Amit I (2016). Dissecting Immune Circuits by Linking CRISPR-Pooled Screens with Single-Cell RNA-Seq. *Cell* 167, 1883–1896.e15. [PubMed: 27984734]
- Kannangai R, Vivekanandan P, Martinez-Murillo F, Choti M, and Torbenson M (2007). Fibrolamellar carcinomas show overexpression of genes in the RAS, MAPK, PIK3, and xenobiotic degradation pathways. *Hum. Pathol* 38, 639–644. [PubMed: 17367606]
- Kinsey C, Balakrishnan V, O’Dell MR, Huang JL, Newman L, Whitney-Miller CL, Hezel AF, and Land H (2014). Plac8 links oncogenic mutations to regulation of autophagy and is critical to pancreatic cancer progression. *Cell Rep* 7, 1143–1155. [PubMed: 24794439]
- Land H, Parada LF, and Weinberg RA (1983). Tumorigenic conversion of primary embryo fibroblasts requires at least two cooperating oncogenes. *Nature* 304, 596–602. [PubMed: 6308472]
- Langfelder P, and Horvath S (2008). WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* 9, 559. [PubMed: 19114008]

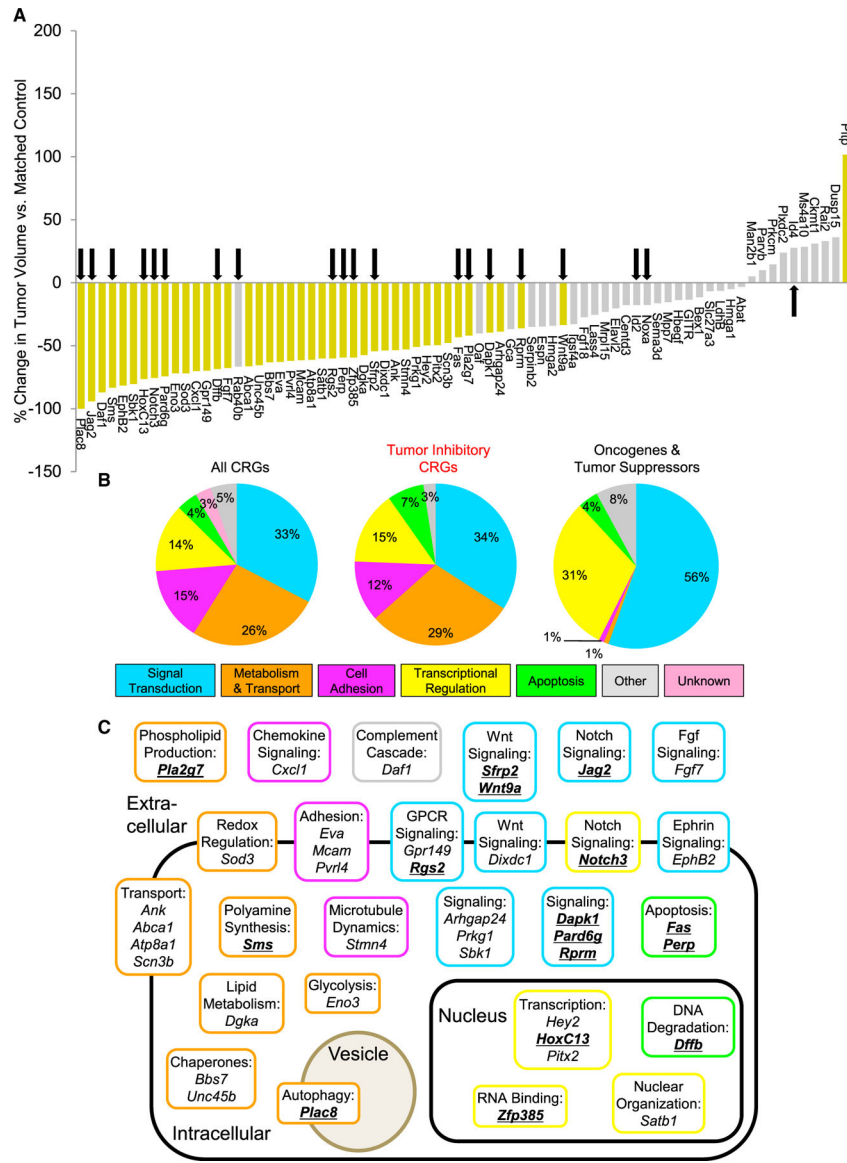
- Lasorella A, Uo T, and Iavarone A (2001). Id proteins at the cross-road of development and cancer. *Oncogene* 20, 8326–8333. [PubMed: 11840325]
- Lloyd AC, Obermüller F, Staddon S, Barth CF, McMahon M, and Land H (1997). Cooperating oncogenes converge to regulate cyclin/cdk complexes. *Genes Dev* 11, 663–677. [PubMed: 9119230]
- Margolin AA, Nemenman I, Basso K, Wiggins C, Stolovitzky G, Dalla Favera R, and Califano A (2006). ARACNE: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. *BMC Bioinformatics* 7 (Suppl 1), S7.
- Markowitz F, and Spang R (2003). Evaluating the effect of perturbations in reconstructing network topologies. In *Proceedings of the 3rd International Workshop on Distributed Statistical Computing (DSC 2003) March 20–22, Vienna., Austria., Hornik K, Leisch F, and Zeileis A, eds.. <http://www.markowitzlab.org/docs/markowitz@dsc2003.pdf>.*
- McCall MN, Uppal K, Jaffee HA, Zilliox MJ, and Irizarry RA (2011). The Gene Expression Barcode: leveraging public data repositories to begin cataloging the human and murine transcriptomes. *Nucleic Acids Res* 39, D1011–D1015. [PubMed: 21177656]
- McCall MN, McMurray HR, Land H, and Almudevar A (2014). On non-detects in qPCR data. *Bioinformatics* 30, 2310–2316. [PubMed: 24764462]
- McMurray HR, Sampson ER, Compitello G, Kinsey C, Newman L, Smith B, Chen SR, Klebanov L, Salzman P, Yakovlev A, and Land H (2008). Synergistic response to oncogenic mutations defines gene class critical to cancer phenotype. *Nature* 453, 1112–1116. [PubMed: 18500333]
- Morgenstern JP, and Land H (1990). A series of mammalian expression vectors and characterisation of their expression of a reporter gene in stably and transiently transfected cells. *Nucleic Acids Res* 18, 1068. [PubMed: 2156225]
- Parmigiani G, Garrett ES, Anbazhagan R, and Gabrielson E (2002). A statistical framework for expression-based molecular classification in cancer. *J. R. Stat. Soc. B* 64, 717–736.
- Pe'er D, Regev A, Elidan G, and Friedman N (2001). Inferring subnetworks from perturbed expression profiles. *Bioinformatics* 17 (Suppl 1), S215–S224. [PubMed: 11473012]
- Pegg AE, and Michael AJ (2010). Spermine synthase. *Cell. Mol. Life Sci* 67, 113–121. [PubMed: 19859664]
- Rattner A, Hsieh JC, Smallwood PM, Gilbert DJ, Copeland NG, Jenkins NA, and Nathans J (1997). A family of secreted proteins contains homology to the cysteine-rich ligand-binding domain of frizzled receptors. *Proc. Natl. Acad. Sci. USA* 94, 2859–2863. [PubMed: 9096311]
- Scharpf R, Garrett ES, Hu J, and Parmigiani G (2003). Statistical modeling and visualization of molecular profiles in cancer. *Biotechniques Mar (Suppl)*, 22–29.
- Schraivogel D, Gschwind AR, Milbank JH, Leonce DR, Jakob P, Mathur L, Korbel JO, Merten CA, Velten L, and Steinmetz LM (2020). Targeted Perturb-seq enables genome-scale genetic screens in single cells. *Nat. Methods* 17, 629–635. [PubMed: 32483332]
- Sewing A, Wiseman B, Lloyd AC, and Land H (1997). High-intensity Raf signal causes cell cycle arrest mediated by p21Cip1. *Mol. Cell. Biol* 17, 5588–5597. [PubMed: 9271434]
- Smith B, and Land H (2012). Anticancer activity of the cholesterol exporter ABCA1 gene. *Cell Rep* 2, 580–590. [PubMed: 22981231]
- Smith B, Schafer XL, Ambeskovic A, Spencer CM, Land H, and Munger J (2016). Addiction to Coupling of the Warburg Effect with Glutamine Catabolism in Cancer Cells. *Cell Rep* 17, 821–836. [PubMed: 27732857]
- Swendsen RH, and Wang JS (1986). Replica Monte Carlo simulation of spin glasses. *Phys. Rev. Lett* 57, 2607–2609. [PubMed: 10033814]
- van de Haar J, Canisius S, Yu MK, Voest EE, Wessels LFA, and Ideker T (2019). Identifying Epistasis in Cancer Genomes: A Delicate Affair. *Cell* 177, 1375–1383. [PubMed: 31150618]
- Werhli AV, Grzegorzczak M, and Husmeier D (2006). Comparative evaluation of reverse engineering gene regulatory networks with relevance networks, graphical gaussian models and bayesian networks. *Bioinformatics* 22, 2523–2531. [PubMed: 16844710]
- Whitehead RH, VanEeden PE, Noble MD, Ataliotis P, and Jat PS (1993). Establishment of conditionally immortalized epithelial cell lines from both colon and small intestine of adult H-2Kb-tsA58 transgenic mice. *Proc. Natl. Acad. Sci. USA* 90, 587–591. [PubMed: 7678459]

- Xia M, and Land H (2007). Tumor suppressor p53 restricts Ras stimulation of RhoA and cancer cell motility. *Nat. Struct. Mol. Biol* 14, 215–223. [PubMed: 17310253]
- Yuan B, Latek R, Hossbach M, Tuschl T, and Lewitter F (2004). siRNA Selection Server: an automated siRNA oligonucleotide prediction server. *Nucleic Acids Res* 32, W130–W134. [PubMed: 15215365]
- Zak DE, Gonye GE, Schwaber JS, and Doyle FJ 3rd. (2003). Importance of input perturbations and stochastic gene expression in the reverse engineering of genetic regulatory networks: insights from an identifiability analysis of an in silico network. *Genome Res* 13, 2396–2405. [PubMed: 14597654]
- Zhang B, and Horvath S (2005). A general framework for weighted gene co-expression network analysis. *Stat. Appl. Genet. Mol. Biol* 4, Article17.
- Zhang Y, Cuerdo J, Halushka MK, and McCall MN (2021). The effect of tissue composition on gene co-expression. *Brief. Bioinform* 22, 127–139. [PubMed: 31813949]
- Zilliox MJ, and Irizarry RA (2007). A gene expression bar code for microarray data. *Nat. Methods* 4, 911–913. [PubMed: 17906632]

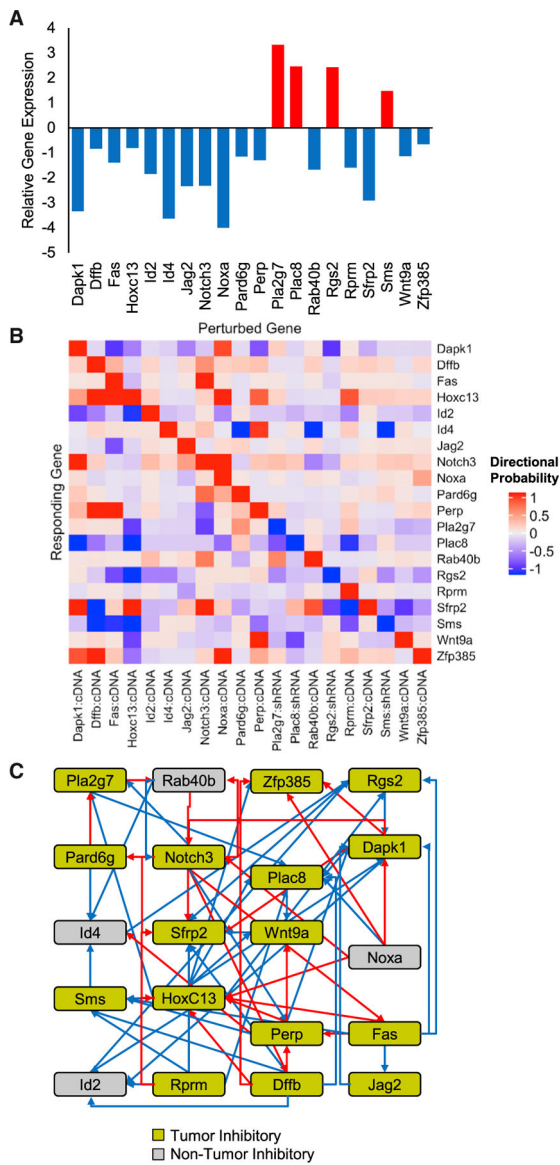
**Highlights**

- A network of non-mutated genes is critical to the malignant state
- TopNet can accurately model cellular responses to genetic perturbations
- TopNet is capable of pinpointing key architectural features of cancer cells
- TopNet has utility for identification of non-mutant targets for cancer intervention





**Figure 1. Impact of CRG perturbations on tumor formation**  
 (A) Waterfall plot indicates percentage change in endpoint tumor volume of allografts following perturbation of individual CRGs in mp53/Ras cells. Perturbations significantly altering tumor size, as compared to matched controls are shown in gold (p < 0.1, Wilcoxon signed-rank test with Benjamini-Hochberg adjustment). Others are shown in gray. Arrows indicate genes chosen for inclusion in gene regulatory network modeling.  
 (B) Pie charts indicate the proportions of functional gene annotations according to the Gene Ontology database for all CRGs, tumor-inhibitory CRGs, and classical oncogenes and tumor suppressors, respectively. Colors signify biological processes, as indicated.  
 (C) Scheme summarizes the cellular localization and cell biological functions of proteins encoded by CRGs, colored as in (B). CRGs chosen for inclusion in gene regulatory network analysis are indicated by bold, underlined text.

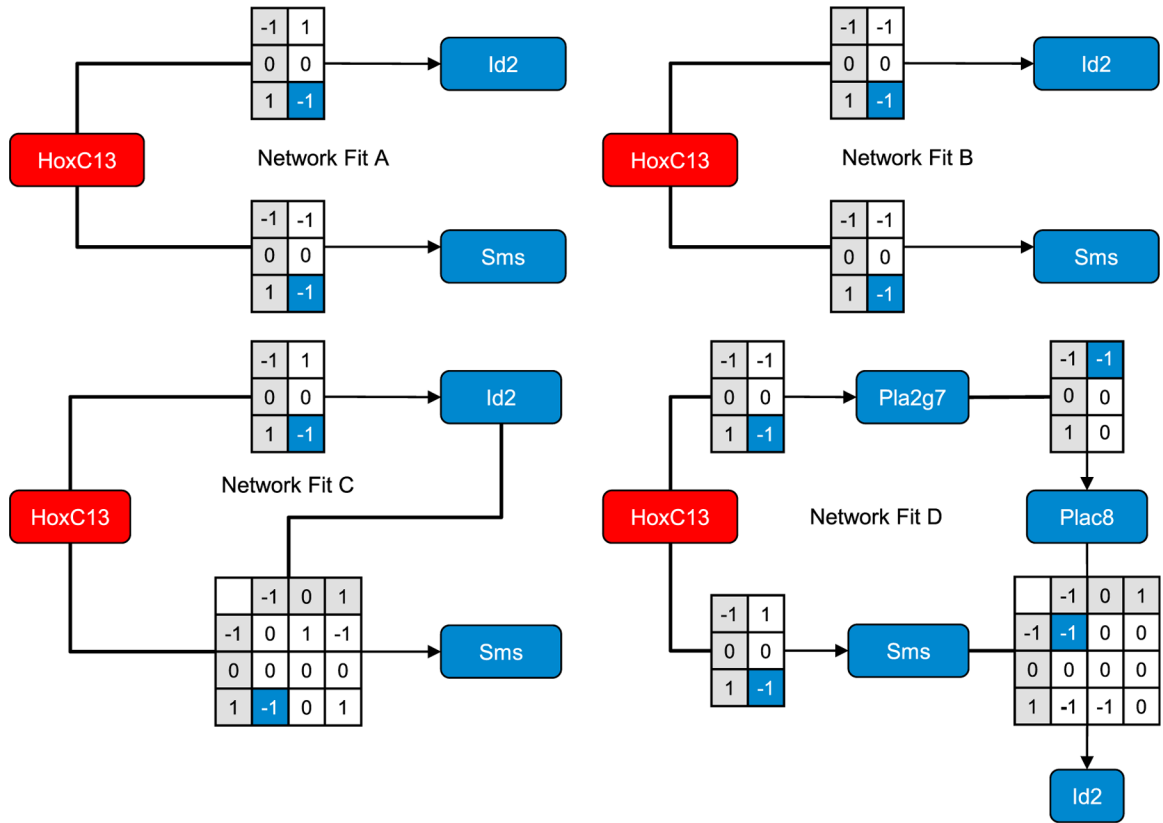


**Figure 2. CRG connectivity**

(A) Bar graph showing the expression levels of each CRG selected for network analysis, as compared to their baseline expression in non-transformed parental YAMC cells. Genes upregulated in mp53/Ras versus YAMC are shown in red, while downregulated CRGs are shown in blue. Measurements were made by TaqMan Low-Density Array (TLDA) for multiple replicates of each gene perturbation.

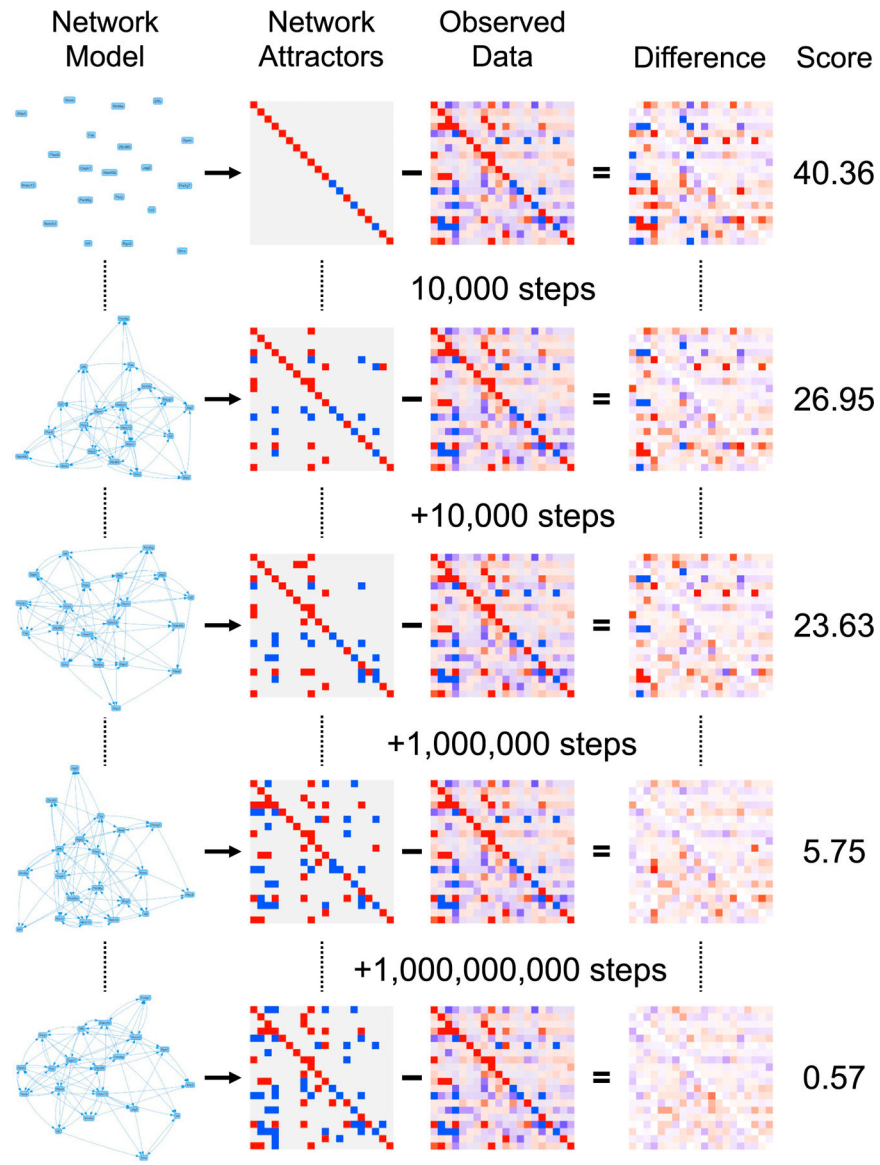
(B) Probability of change in expression of each CRG in response to individual CRG perturbations indicated, computed based on mRNA levels detected on TLDA for multiple replicates of each gene perturbation. Red denotes a likely increase in the expression of the downstream gene and blue denotes a likely decrease in expression. Note that the diagonal indicates the direction of perturbation, intended to restore levels of gene expression found in parental YAMC cells for each perturbed gene.

(C) Connectivity graph showing the influence of given CRGs on the expression of other CRGs. Arrows indicate parent-child connections with  $|\text{probability}| > 0.5$ , colored as in (A) to indicate likely direction of change in the child. Nodes are colored as in Figure 1 to indicate tumor inhibitory effects of perturbation of the indicated CRGs (gold = inhibitory, gray = no significant change). Placement of nodes was chosen for comparison to Figure 5.



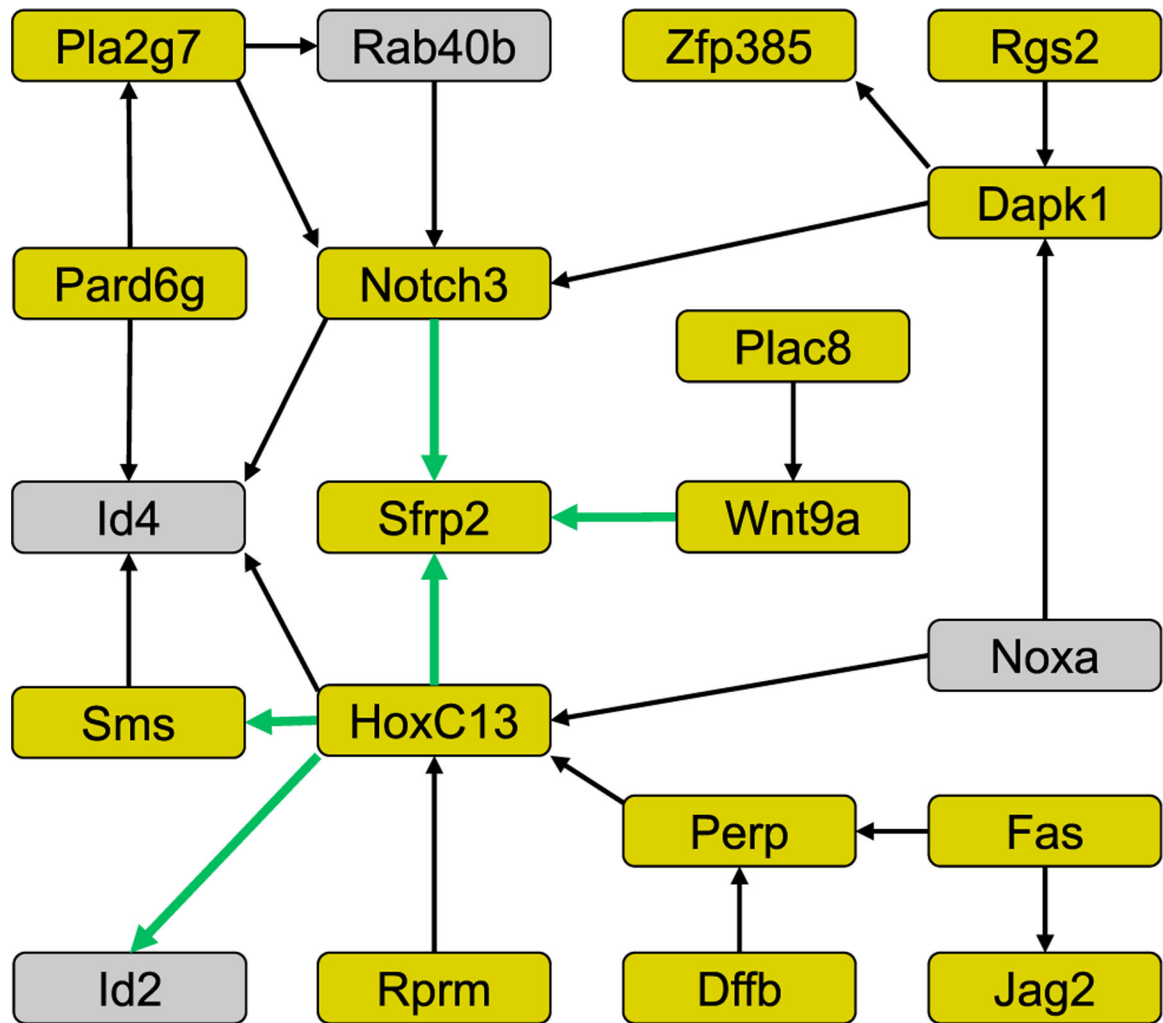
**Figure 3. Transition functions encode network behavior**

Paths from perturbation of parent node HoxC13 to its possible children are diagrammed. Four network fits of 100 computed are represented. Marginal effects (single parent) and joint effects (multiple parents) are shown with transition function tables. For simplicity of visualization, the maximum in-degree shown here is 2, but the models applied allow for up to 4 parents to influence a given child node's expression. Red nodes indicate that the gene increases its expression, while blue denotes reduced expression.



**Figure 4. Gene regulatory network reconstruction process**

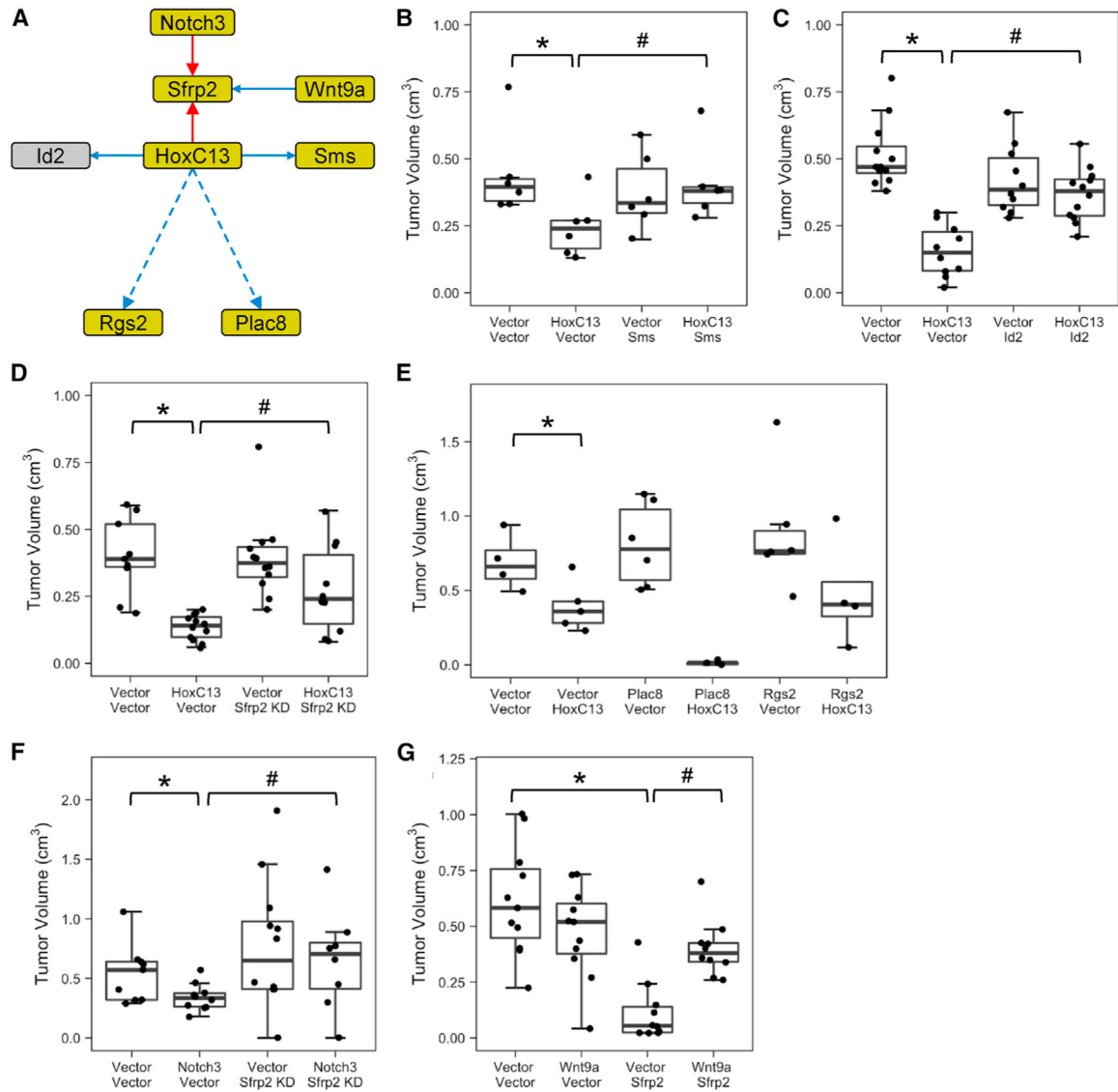
The network modeling algorithm begins with a null network in which no gene affects any other gene (top left). This produces a set of null attractors wherein the perturbation of any gene does not produce a change in any other gene. The network models are scored by summing the absolute value of the differences between the probabilities of up- or downregulation and the attractors, which are then standardized by subtracting the best theoretically possible score. Comparing the null attractors to the calculated probabilities results in substantial differences, which produce a score of 40.36, indicating that the network model does not explain the observed data well. Random changes in topology and/or transition functions are made to the network model and the process is repeated. Here, we show snapshots of the process after the first 10,000 steps, the next 10,000 steps, then after an additional 1 million steps, and finally after 1 billion more steps, at which point, the network model produces attractors that match the probabilities calculated from the data very well.



**Figure 5. Summary of CRG network topology**

The network models produced by 100 independent fits are mapped by reporting the proportion of networks in which an edge (parent-child relationship) appears in >80% of the network models. The node color indicates the effect of perturbation of that node on tumor formation capacity (as shown in Figure 1 and used in Figure 2, gold = tumor inhibitory perturbation, gray = no significant change). Experimentally tested epistatic interactions are indicated by green arrows.





**Figure 6. Epistatic CRG interactions affecting cancer phenotype**

(A) Diagram of a subnetwork of the GRN shown in Figure 5, whose edges are indicated by solid lines, with the addition of 2 edges seen only in the connectivity graph but not in the topological map (dashed lines). Red arrows indicate that perturbation of the parent gene increases the expression of the child, while blue arrows denote reduced expression of the child in response to the parent perturbation. Nodes are colored as in Figures 2 and 5. (B–G) Boxplots show tumor growth in response to indicated single and double perturbations targeting parent node HoxC13 and its predicted children, or child node Sfrp2 and its parents. KD denotes knockdown perturbations; all other perturbations induce overexpression. Perturbations significantly decreasing tumor size, as compared to matched controls are indicated by an asterisk (\* $p < 0.05$ , unadjusted Wilcoxon signed-rank test). Double gene perturbations showing significantly different tumor size as compared to individual gene

perturbations are indicated by a hash mark (#,  $p < 0.1$ , unadjusted Wilcoxon signed-rank test versus each individual perturbation).

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

## KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
<b>Bacterial and virus strains</b>		
pBabe retroviral vector containing either puromycin, blasticidin or hygromycin resistance genes	Addgene	pBabe-puro: pBabe-hygro: pBabe-blast
pSuper retroviral vector containing puromycin resistance gene	Oligoengine	pSuper.retro.puro
pLKO lentiviral vector containing puromycin resistance gene	Addgene	pLKO.1 puro
pLenti/Ubc/V5 lentiviral vector containing puromycin resistance gene	Invitrogen	pLenti6/Ubc/V5
<b>Chemicals, peptides, and recombinant proteins</b>		
RPMI 1640 media	Invitrogen	11875119
Fetal bovine serum	Hyclone	SH30071
ITS-A	Invitrogen	51300044
Gentamycin	Invitrogen	15750060
Interferon gamma	R&D Systems	485MI
Rat tail derived collagen IV	Corning	354236
Polybrene	Sigma-Aldrich	H9268
Puromycin	Sigma-Aldrich	P7255
Hygromycin	Invitrogen	10687010
Blasticidin	Invitrogen	R21001
<b>Critical commercial assays</b>		
Please see Table S3 for listing of TaqMan qPCR assays utilized in TaqMan Low Density Array (TLDA) assays	Applied Biosystems	N/A
RNeasy Mini Kit with on-column DNase digestion	QIAGEN	74106
High Capacity cDNA Reverse Transcription Kit	Applied Biosystems	4368814
iQ SYBR Green qPCR Master Mix	Bio-Rad	1708884
<b>Experimental models: cell lines</b>		
Young adult mouse colon cells derived from the H-2Kb / tsA58 transgenic mouse	Gift of R. Whitehead and A.W. Burgess	YAMC cells
“Phoenix” cells producing murine ecotropic virus (ΦNX-ECO or Phoenix-ECO)	ATCC	Phoenix-ECO
Human embryonic kidney 293 cells expressing SV40 T antigen	ATCC	HEK293T
<b>Experimental models: organisms/strains</b>		
CrI: CD-1-Foxn1 <sup>tm</sup> female mice	Charles River Laboratories	CrI: CD-1-Foxn1 <sup>tm</sup>
<b>Oligonucleotides</b>		
Please see Table S2 for listing of primers for SYBR-green qPCR	IDT	N/A
<b>Recombinant DNA</b>		
Please see Table S1 for listing of plasmids generated for these studies		
<b>Software and algorithms</b>		
Sequence Detection System (SDS) software, versions 2.0 and 3.0	Applied Biosystems	N/A
Non-detect imputation algorithm (nondetects package)	Bioconductor	<a href="https://doi.org/10.18129/B9.bioc.nondetects">https://doi.org/10.18129/B9.bioc.nondetects</a>

REAGENT or RESOURCE	SOURCE	IDENTIFIER
TopNet algorithm (ternarynet package)	Bioconductor	<a href="https://doi.org/10.18129/B9.bioc.ternarynet">https://doi.org/10.18129/B9.bioc.ternarynet</a>
Reproducible workflow for this paper (crgnet package)	This paper	Methods S1

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript