

RESEARCH ARTICLE

Open Access



High-throughput validation of ceRNA regulatory networks

Hua-Sheng Chiu¹, María Rodríguez Martínez², Mukesh Bansal³, Aravind Subramanian⁴, Todd R. Golub^{4,5,6}, Xuerui Yang^{7*}, Pavel Sumazin^{1*} and Andrea Califano^{3,8,9*}

Abstract

Background: MicroRNAs (miRNAs) play multiple roles in tumor biology. Interestingly, reports from multiple groups suggest that miRNA targets may be coupled through competitive stoichiometric sequestration. Specifically, computational models predicted and experimental assays confirmed that miRNA activity is dependent on miRNA target abundance, and consequently, changes in the abundance of some miRNA targets lead to changes to the regulation and abundance of their other targets. The resulting indirect regulatory influence between miRNA targets resembles competition and has been dubbed competitive endogenous RNA (ceRNA). Recent studies have questioned the physiological relevance of ceRNA interactions, our ability to accurately predict these interactions, and the number of genes that are impacted by ceRNA interactions in specific cellular contexts.

Results: To address these concerns, we reverse engineered ceRNA networks (ceRNETs) in breast and prostate adenocarcinomas using context-specific TCGA profiles, and tested whether ceRNA interactions can predict the effects of RNAi-mediated gene silencing perturbations in PC3 and MCF7 cells. Our results, based on tests of thousands of inferred ceRNA interactions that are predicted to alter hundreds of cancer genes in each of the two tumor contexts, confirmed statistically significant effects for half of the predicted targets.

Conclusions: Our results suggest that the expression of a significant fraction of cancer genes may be regulated by ceRNA interactions in each of the two tumor contexts.

Keywords: ceRNA, microRNA, LINCS, PRAD, BRCA

Background

Identifying regulatory interactions that mediate the effects of genomic alterations is a necessary step for interpreting the function of trans-acting variants in complex diseases, including cancer [1, 2]. Among these, miRNA dysregulation, arising from alterations targeting their transcriptional [3] or biogenesis regulators [4], plays an established role in tumorigenesis [5]. Recently, multiple groups have reported on gene products that modulate miRNA activity, including RNA species that can alter the abundance of other

RNAs *in trans* through ceRNA interactions [6–17]. These studies show that targets of the same miRNAs are coupled, and that up- or down-regulation of one target may alter the expression of other cognate targets by sequestering or releasing their shared miRNA molecules, respectively (Fig. 1a).

Since the discovery of ceRNA regulation in human cells [9, 10] multiple reports questioned the physiological relevance of ceRNA interactions, researcher's ability to predict them, and the number of genes that are affected in each context [18–20]. To address these concerns, we proceeded to test genome-wide ceRNA predictions made by the information-theoretic Hermes algorithm [7]. For the sake of generality, we performed this analysis in two distinct tumor contexts, using a set of large-scale and high-throughput shRNA-mediated perturbation assays in model cell lines assembled by the Library of Integrated Network-based Cellular Signatures (LINCS) [21]. Specifically, we inferred ceRNETs using TCGA profiles of prostate

* Correspondence: yangxuerui@tsinghua.edu.cn; sumazin@bcm.edu; ac2248@cumc.columbia.edu

⁷MOE Key Laboratory of Bioinformatics, Tsinghua-Peking Center for Life Sciences, School of Life Sciences, Tsinghua University, Beijing 100084, China
¹Texas Children's Cancer Center and Department of Pediatrics, Baylor College of Medicine, Houston, TX, USA

³Columbia Department of Systems Biology, Center for Computational Biology and Bioinformatics, Herbert Irving Comprehensive Cancer Center, Columbia University, New York, NY10032USA

Full list of author information is available at the end of the article



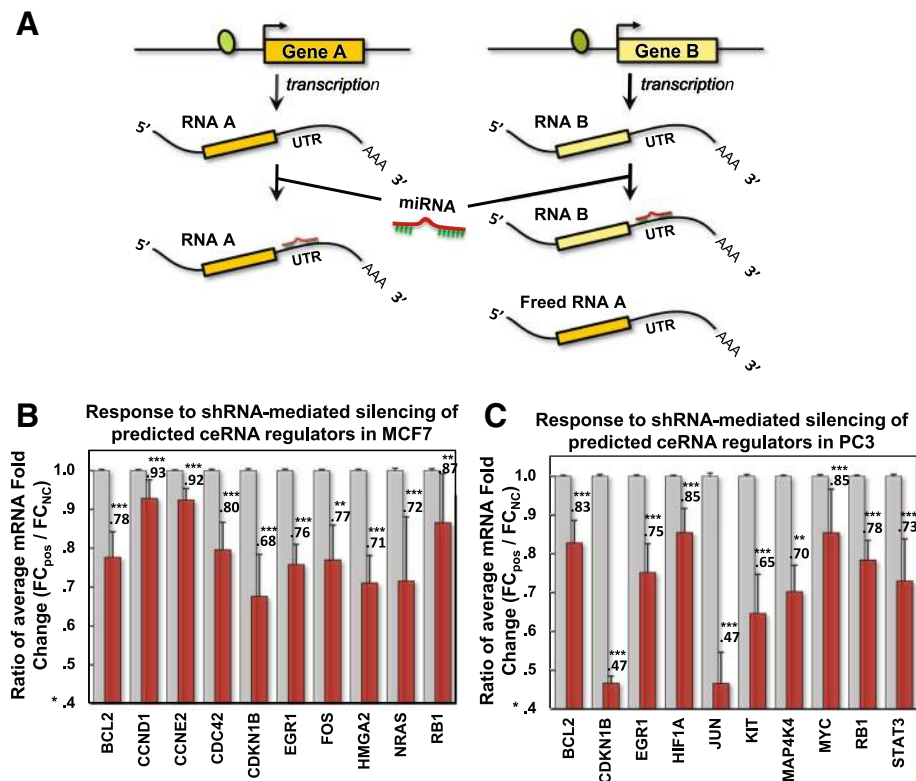


Fig. 1 Model and validation of miRNA-target coupling. **a** RNAs up and down regulate one another by titrating shared miRNA regulators. Up regulation of RNA B sequesters shared miRNAs, leading to weaker miRNA-mediated repression of RNA A transcripts and its consequent up regulation. **b** In order to validate predicted interaction networks on a large scale, we evaluated whether interactions are predictive of global mRNA expression changes following shRNA perturbations using LINCSeq. A selection of known cancer genes in breast cancer and **c** prostate adenocarcinomas were effectively repressed following silencing of their predicted ceRNA regulators in MCF7 and PC3, respectively. Red bars represent average fold changes of a target ceRNA relative to non-targeting controls (gray bars) following silencing of its predicted ceRNA regulators at select time points; see Figs. 2 and 3 for details. Data are represented as mean ± SEM

and breast adenocarcinomas and tested them using a LINCSeq compendium of perturbation profiles, representative of the shRNA-mediated silencing of >3000 genes in PC3 and MCF7 cells [21–23]. We propose that the high validation rates of these assays can inform on the accuracy of computational predictions, and will help estimate the number of genes that are modulated by ceRNA in representative tumor contexts.

Methods

We begin by describing our ceRNA-prediction method, followed by a description of LINCSeq data used to test the predictive ability of inferred ceRNA interactions in PC3 and MCF7 cells. We then describe our efforts to test whether systematic biases affect our test results.

Inference method

We used an extended version of the Hermes algorithm, which we had previously introduced to discover glioma-specific ceRNAs [7], to systematically discover ceRNA interactions in prostate (PRAD) and breast (BRCA)

adenocarcinomas, using matched miRNA and mRNA expression profiles of the corresponding TCGA cohorts. While the ceRNA inference component of the algorithm was unchanged, the new algorithm also supports the identification of the specific miRNAs that mediate each interaction (*mediators*); these miRNAs are predicted to target both mRNAs in a ceRNA interaction, and their activity is affected (modulated) by target mRNA abundance. This extension is useful for generating more specific hypotheses for future functional testing.

We note that Hermes-inferred ceRNA interactions are independent of the co-expression of coupled mRNAs; rather, they are based on assessing whether the abundance of one mRNA species modulates abundance of the other (and vice-versa), via their shared miRNA program. This assessment is based on the statistical significance of the mutual information between the abundance of one mRNA species and one or more miRNA, given the abundance of another mRNA targeted by the same miRNAs. We note that a majority of predicted ceRNA interactions involved mRNAs that are not significantly co-expressed, and co-expression

did not implicate genes as ceRNA interacting partners. This is contrasted with MuTaME [11], which uses co-expression between ceRNAs to predict interactions, and cefinder [24], which is based on sequence analysis alone.

Hermes predicts ceRNA-coupling between two mRNAs based on the relative size of their shared miRNA regulatory program, as predicted by the Cupid algorithm [7], and the conditional mutual information between one of the mRNAs and each of their shared miRNAs, given the other mRNA. Namely, given genes T_i and T_j , and the set of miRNAs that regulate them $\Pi_{miR}(T_i)$ and $\Pi_{miR}(T_j)$, their shared program is identified by taking the intersection $\Pi_{miR}(T_i; T_j) = \Pi_{miR}(T_i) \cap \Pi_{miR}(T_j)$. First, Hermes tests that the size of $\Pi_{miR}(T_i; T_j)$ relative to the sizes of the individual programs is statistically significant at $FDR < 0.01$ by Fisher's exact test (FET). Then, Hermes evaluates the statistical significance p_{kij} (p -value) of the test $CMI[miR_k; T_i | T_j] > MI[miR_k; T_i]$, where CMI and MI stand for conditional mutual information and mutual information respectively, and the variables indicate the expression of the corresponding RNA species [7].

The CMI is estimated using an adaptive partitioning algorithm [25] by first iteratively partitioning the 3-dimensional expression space evenly into 8 partitions per iteration until partitions are balanced ($p > 0.05$ by Chi-squared test), and then summing up CMI across partitions. P -values for each triplet are computed based on a null-hypothesis where the candidate modulator's expression (T_i) is shuffled 1000 times, thus preserving the pairwise mutual information between miRNA and target. Final significance across the entire set of miRNA mediators is computed using Fisher's method to integrate both regulatory directions, i.e. T_i affecting miR_k regulation of T_j as well as T_j affecting miR_k regulation of T_i , for all miRNA mediators $\Pi_{miR}(T_i; T_j)$. Specifically, $X^2 = -2 \sum_{k=1}^N \ln(p_{kij}, p_{kji})$ is computed and used to estimate a significance p -value for the entire program. Note that X^2 follows a Chi-square distribution, with $4N$ degrees of freedom, where N is the number of miRNAs in the shared program. Finally, only predictions passing significance of $FDR < 1E-3$ are selected. Note that selected predictions by Hermes have been previously validated in glioblastoma cell lines [7]. In addition, the presence of transcripts with alternative 3' UTRs is expected to reduce the sensitivity of prediction.

In order to identify miRNA mediators in addition to ceRNA interactions, we modified Hermes to perform greedy addition of miRNA mediators and to optimize the combined p -value for each predicted interaction. Namely, for each candidate interaction, we searched for the minimum combined p -value through the greedy forward inclusion of individual miRNAs. Additional miRNAs were included as candidate mediators only if they improve the

joint p -value, as estimated using Fisher's method. MiRNAs failing to improve the joint p -value lack functional evidence for mediating the interaction and were thus excluded from the analysis.

The LINCS database and its analysis

The LINCS database includes Luminex-based multiplexed assays to measure the expression of 1171 genes (L1000) in response to a variety of perturbations. Selected perturbations include shRNA-mediated silencing (in triplicate) of 1845 genes participating in ceRNA interactions, in both BRCA (MCF7) and PRAD (PC3) cell lines (Additional file 1: Tables S3 and S4). Gene expression was measured using the L1000 assay at two time points (96 and 144 h), following each perturbation. To achieve adequate statistical power, we limited our tests to genes with six or more silenced Hermes-inferred ceRNA regulators. In total, we evaluated predicted ceRNA targeting of 405 genes (of which 365 were validated at 96 h and 398 at 144 h) in MCF7 cells and 419 genes (of which 363 were validated at 96 h and 376 at 144 h) in PC3 cells.

In LINCS perturbation assays, while some data points are missing due to quality control metrics, the expression of most genes was profiled in triplicates at both 96 and 144 h after shRNA transduction. On average, 3.3 unique shRNA hairpins were used to silence each of 1845 breast and prostate oncogenes and tumor suppressors (*cancer genes*) in our networks. By definition, for each interaction, the ceRNA regulator is the one targeted by the shRNA perturbation and the ceRNA target is the one profiled after silencing to determine any interaction mediated change. Fold-change for each target, in response to the silencing of one of its regulators, was estimated by averaging across all shRNA hairpins targeting this regulator. The identities of the specific shRNA hairpins, regulators, and targets are provided in Additional file 1: Table S5.

In total, 9055 and 9800 predicted BRCA interactions were tested in MCF7 cells at 96 and 144 h after shRNA-mediated silencing, respectively. Similarly, 8858 and 10,213 predicted PRAD interactions were tested in PC3 cells at 96 and 144 h after silencing, respectively. Due to the small number of replicates, it is not possible to evaluate the statistical significance of individual predicted interactions. Instead, we evaluated the average effects of all ceRNA regulators in the list of silenced genes on the expression of a given ceRNA target. Average fold changes and associated standard errors were computed by comparing to non-targeting controls, at each time point, and in each relevant cellular context. Additional file 1: Tables S3 and S4 provide mRNA expression fold change measurements following shRNA-mediated perturbation of breast and prostate cancer genes in MCF7 and PC3, respectively, based on the LINCS perturbation assays.

The size of this dataset allows for rigorous controls that avoid bias due to gene expression variability and off-target effects of RNAi-mediated perturbations. To estimate the statistical significance of target responses to gene perturbations, we examined fold change in target expression following silencing of its predicted regulator, as well as the effect of silencing a specific regulator across all of its inferred targets. For the latter, we associated each candidate target with a rank vector representing its percentile ranked relative fold-change at a specific time point following each gene silencing perturbation. We then ranked expression fold-changes of all profiled genes at a specific time point following silencing of any given gene and compared ranks across different perturbations. In this way, each candidate target was associated with a rank vector, representing its relative fold-change following each perturbation. Considering each target in isolation, we compared its response to perturbations of its predicted regulators as well as its response to shRNA-mediated silencing of other genes. Mann–Whitney U test was used to determine whether the rank of a target following silencing of its predicted regulators was lower than that following silencing of other genes.

Specifically, Fold change measurements for up to 1171 genes in response to a given perturbation allowed ranking of the profiled genes based on the strength of the response. First, we assigned significance to the response of a gene to the silencing of its predicted regulators by comparing the set of its scores associated with perturbations of predicted regulators to the scores of all other genes, i.e. the gene's ranks following silencing of its predicted regulators vs. its ranks following silencing of all other genes. Then we used a Mann–Whitney test to determine whether the ranks of a target after silencing of its regulators was significantly lower than its ranks following silencing of all other genes. Given the number of gene perturbations (up to 1171 gene silencing experiments), the two sets of ranks were expected to be normally distributed and can be approximated by a z-score and a corresponding p value. On average, for each target gene, the number of perturbations targeting its regulators was 1% of the total number of perturbations tested.

Standard error was computed using standard error propagation techniques, i.e., the standard error was estimated as

$\sqrt{\sum_{i=1}^N \sigma_i^2}$, where σ_i is the standard deviation of each individual silencing experiment and N is the number of tested regulators. When comparing responses of ceRNA targets to controls, as described in Fig. 1b-c, we randomly assembled control sets composed of as many non-ceRNA regulators as the number of ceRNA regulators for each target. We then calculated the average fold change and the error-propagated standard error after silencing each of non-ceRNA regulator and estimated the significance of fold changes using a two-tailed rank-sum

test. This process was repeated 1000 times to obtain averaged fold changes, propagated standard errors, and averaged p values based on the negative controls.

Accounting for systematic biases in LINCS analysis

To ensure that comparisons of target responses to predicted regulators and non-regulators are free of bias, we repeated the analysis presented for MCF7 and PC3 after controlling several properties. These included re-analyses after (1) eliminating outlier responders, i.e. ceRNA interactions associated with the most significant ceRNA-validating responses; (2) eliminating shRNAs that can act as human miRNAs [26] and produce off-target effects; and after accounting for (3) 3' UTR length and (4) CG content, (5) RMA-normalized expression, (6) expression variability, and (7) expression correlation with the predicted target. Results are presented in Additional file 2: Figure S1 and suggest that accounting for these potential biases had relatively little effect on the number of targets that were found to be significantly downregulated by silencing their predicted regulators. We note that the average gene expression change across all interactions in LINCS is 1.0.

When eliminating outliers we removed predicted ceRNA interactions where the inducing effects were greater than $Q1-1.5*STD$ by percentile rank from the analysis. By discarding the strongest ceRNA-like effects, we eliminated any chance that the test may be biased by a few outlier events. To eliminate potential off-target effects caused by miRNA-like behavior, we eliminated all shRNAs whose 7-base seed subsequence (2nd to 8th position) matched miRBase human miRNA seeds. To study the effects 3' UTR length and composition, we binned all potential ceRNA regulators including predicted regulators and controls; 3' UTRs were binned by either length using 25-base offsets or by GC content in 0.001 intervals. Length and content were studied independently. When comparing ceRNA interactions to non-interactions (controls), both were taken from corresponding bins. To study the effects of expression magnitude, we averaged MCF7 and PC3 gene RMA-normalized expression across 213 MCF7 and 64 PC3 experiments deposited in Gene Expression Atlas [27], binned at 0.01 intervals. Expression variability describes median absolute deviation, binned at 0.001 intervals. Expression correlation was measured using Spearman correlation with the expression profile of the target, binned at 0.01 intervals.

Integrative Statistical Evaluation

To evaluate the significance of all tested interactions at each time point in both cell lines, we used a one-sample

Kolmogorov-Smirnov test [28]. This normality test evaluates whether z scores obtained from fold-change rank comparisons follow a standard normal distribution with $\mu = 0$ and $\sigma = 1$, thus assigning significance against the null hypothesis that z scores are selected at random. We ranked targets based on z scores and calculated the expected p value when assuming that z scores were drawn from a standard normal distribution. The result is p -value estimates for each time point for each of the two cell lines, taken in aggregate across all tested interactions. We did not aggregate across time points and cell lines to avoid statistical dependence that is sure to result from using the same shRNAs in multiple assays.

Results

We first describe ceRNA regulatory networks that were inferred using miRNA and mRNA expression profiles of breast and prostate adenocarcinomas. These networks were used to predict secondary effects by shRNAs that target predicted ceRNA regulators of cancer genes in both contexts. Taken as a whole, we then evaluated whether ceRNA interactions can predict the effects of multiple shRNAs on cancer genes. Having demonstrated that ceRNA interactions are predictive of shRNA effects based on LINCS data, we finally compared the predictive abilities of ceRNA interactions predicted by Hermes, MuTaME, and cefinder.

Inferred ceRNETS

We used Hermes (see Methods) to construct ceRNETS using matched TCGA gene and miRNA expression profiles of breast (BRCA) and prostate (PRAD) adenocarcinomas. For PRAD, this included data from 140 samples, representing 23,614 genes and 367 miRNAs [29]; while for BRCA, we used 207 samples, representing 18,748 genes and 524 miRNAs [30]. Predicted ceRNA networks included 476,456 and 447,011 interactions, for the PRAD and BRCA ceRNETS, respectively; see Tables S1-2, where each ceRNA interaction is defined by two RNAs and the miRNAs that couple them. We note that by the time of the article's submission the number of publicly available samples with both miRNA and mRNA expression profiles by TCGA has increased; however, the number of samples we used is sufficient for ceRNA prediction by Hermes.

Due to their size, experimental validation of reverse-engineered networks is often challenging. Consequently, validation is generally performed only on a handful of interactions [31, 32] or on small subnetworks [33, 34]. To validate our inferred interactions on a more realistic scale, we used a large collection of shRNA-mediated silencing assays in the Library of Integrated Network-based Cellular Signatures (LINCS) database [21].

Predicted cancer genes are affected by silencing of their ceRNA regulators

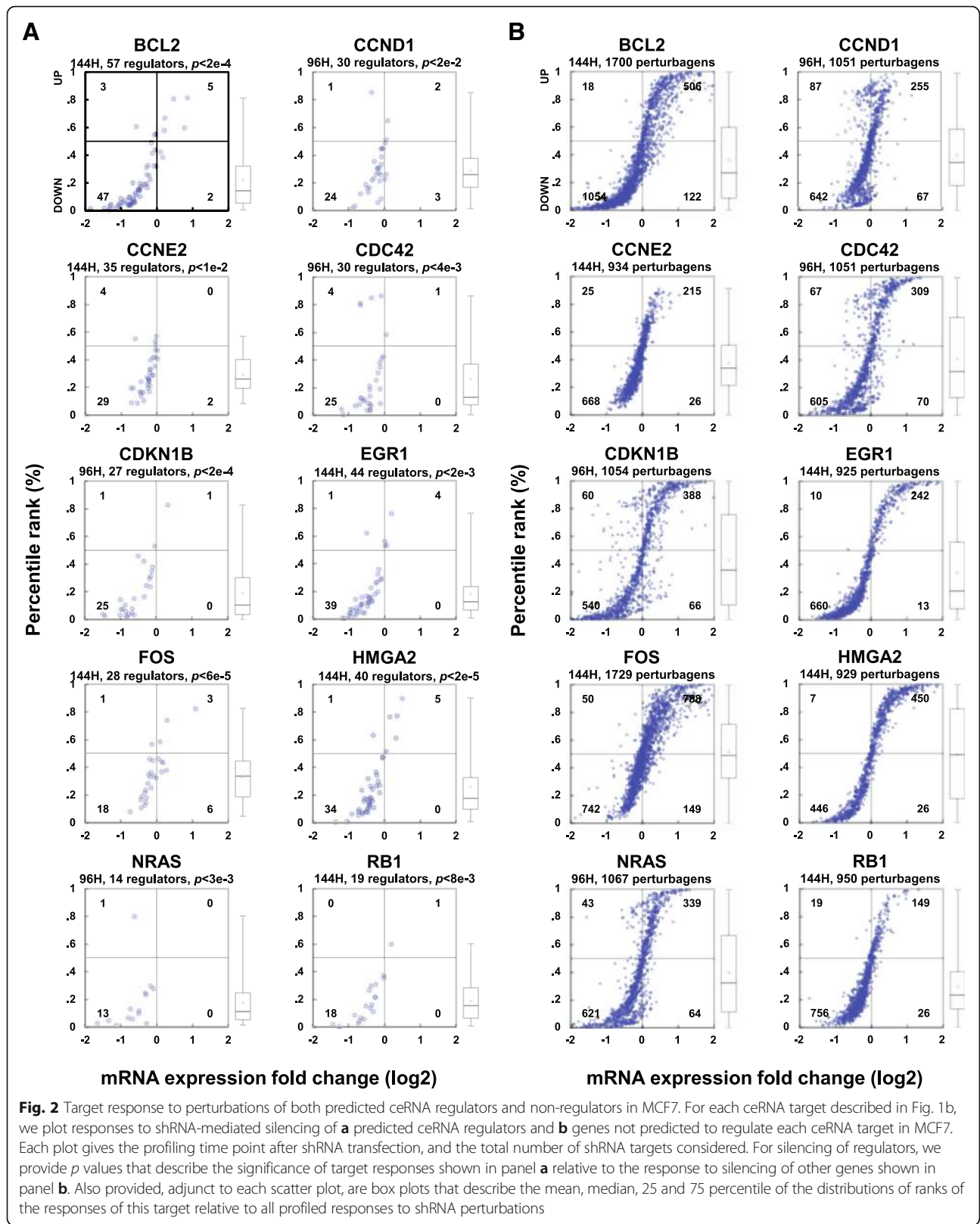
Figure 1b-c describe the average response of a target at a given time point to silencing of all of its tested regulators. Figures 2 and 3 show the responses of targets described in Fig. 1b-c to perturbations of both predicted regulators and non-regulators in MCF7 and PC3. For perturbations of regulators, we provided p values that describe the significance of target responses. An evaluation of all individual interactions and overall responses of each target in both networks is provided in Additional file 1: Table S5.

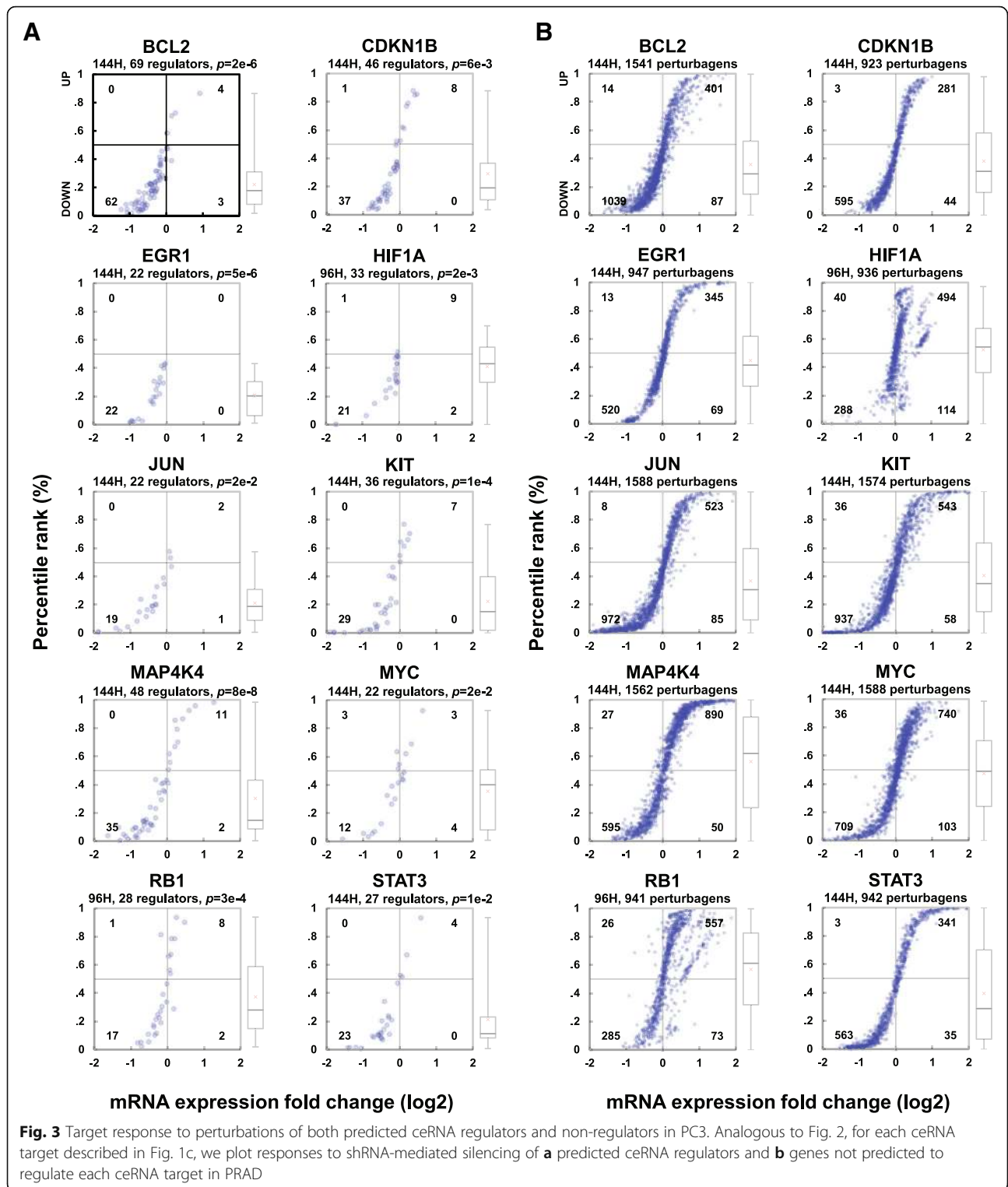
We compared fold changes for every ceRNA target following silencing of its predicted ceRNA regulators to controls, as reported in Methods. Results from these assays confirmed that Hermes predictions are highly enriched in bona fide ceRNA interactions in both BRCA and PRAD and that these interactions may affect the activity of key cancer genes. For instance, 10 established driver cancer genes in BRCA (*BCL2*, *CCND1*, *CCNE2*, *CDC42*, *CDKN1B*, *EGR1*, *FOS*, *HMGA2*, *NRAS* and *RB1*) and PRAD (*BCL2*, *CDKN1B*, *EGR1*, *HIF1A*, *JUN*, *KIT*, *MAP4K4*, *MYC*, *RB1* and *STAT3*) were significantly downregulated when their Hermes-inferred ceRNAs were silenced but were unaffected by silencing of negative control genes (i.e., genes not predicted as their cognate ceRNA regulators); see Figs. 1b-c, 2 and 3.

In total, 69 and 62% of Hermes-inferred targets were significantly down-regulated ($p < 0.05$ by U test) following shRNA-mediated silencing of their Hermes-inferred ceRNAs in MCF7 and PC3, respectively, at least at one time point; see Fig. 4 and Additional file 1: Table S5. Fold-change and p -values were measured by comparing average differential expression of a gene following shRNA-mediated silencing of its inferred ceRNA regulators, compared to silencing of all other genes not predicted to be ceRNA targets of these regulators. This guaranteed the most unbiased selection of negative control assays possible. Moreover, our efforts to control for specific effects that could potentially bias this comparison—including shRNA off-target effects and outliers—reaffirmed analysis results.

Integrative Statistical Evaluation

To evaluate the significance of all tested interactions at each time point in both cell lines, we used a one-sample Kolmogorov-Smirnov test and described in Methods. Our results suggested that ceRNA interactions, even when disregarding all other regulatory modalities, are highly predictive of assay observations: $p < 2e-96$ and $p < 1e-123$ for MCF7 at 96H and 144H, and $p < 6e-84$ and $p < 3e-96$ for PC3 at 96H and 144H, respectively. In MCF7 cells, of the 365 and 398 genes that could be tested at 96 and 144 h





following silencing of their predicted ceRNA regulators, respectively, the vast majority (i.e., 337 and 363) were downregulated. Of these, 181 and 202 were significantly downregulated ($p < 0.05$); only 28 and 35 genes were upregulated, none significantly.

Similarly, for PC3 cells, 319/363 and 336/376 were downregulated (at 96 and 144 h, respectively). Of these, 170 and 174 were significantly downregulated ($p < 0.05$); only 44 and 40 were upregulated, none significantly.

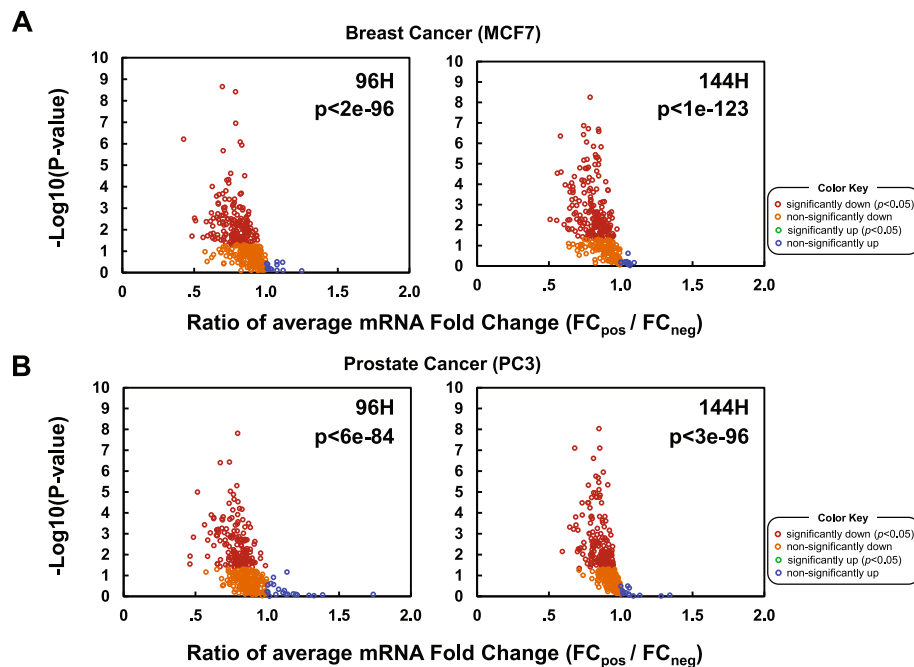


Fig. 4 Statistical evaluation. We plot p -values and average fold changes of target ceRNA expression following silencing of their predicted regulators, compared to silencing of all other genes in both BRCA and PRAD ceRNETS, at two profiling time points in **a** MCF7 and **b** PC3 cells. Results for targets with six or more perturbed ceRNA regulators are shown. To estimate p values for each ceRNA target, we collected all tested regulators and compared average fold-change responses following silencing of inferred ceRNA regulators (FC_{pos}) vs. silencing of all other genes (FC_{neg}) in the network; see Figs. 1, 2 and 3 for illustrative example cancer genes. In total, 91% and 92% (50% and 47% significantly, at $p < 0.05$ by U test) of ceRNA targets, predicted in breast and prostate cancer, were downregulated in response to ceRNA regulator silencing in MCF7 and PC3, respectively. In total, 342 tested ceRNA targets were significantly down-regulated and none were significantly up-regulated. Comparing the number of targets with significantly low FC_{pos} and FC_{neg} fold changes by Mann–Whitney U -test suggests an FDR < 0.01 for overall network validation

In total, 342 tested ceRNA targets were significantly downregulated in at least one assay, at an average of more than 2 assays per target, while none were significantly up-regulated ($p < 0.05$). In aggregate, down-regulation of predicted ceRNA targets was highly significant ($p \leq 6.0E-84$ for each time point and cell line). This analysis constitutes the largest scale validation of a regulatory network to date and suggests that hundreds of cancer genes may be altered through competition for miRNA regulation in BRCA and PRAD. Critically, these results are not merely a reflection of intrinsic coupling of gene expression in cellular systems. Indeed, equivalent numbers of interactions selected at random from non-predicted ceRNAs produced no statistically significant trends. In total, nearly 50% (50% in MCF7 and 47% in PC3) of all predicted targets were significantly downregulated by shRNA-mediated silencing of their predicted ceRNA. We note that when considering individual interactions, 31% of targets were downregulated following silencing of their predicted regulators.

Comparisons with other ceRNA prediction methods

To test whether Hermes predictions are uniquely enriched for down regulation in LINCS data, we used

LINCS assays to test predictions by MuTaME [11] and cefinder [24]; note that MuTaME is context specific and integrates sequence patterns and co-expression between candidate ceRNA to make predictions, while cefinder makes predictions based on sequence information alone. Our results suggest that while Hermes significantly outperformed both methods (Additional file 3: Figure S2), MuTaME and cefinder predictions are significantly enriched in downregulated genes following shRNA-mediated silencing of their regulators.

We used all available predictions by cefinder and MuTaME for our comparison. cefinder scores ceRNA interactions based on the number of miRNA binding sites from the common miRNA program between the ceRNA target X and ceRNA regulator Y ; only the top 50 ceRNA regulators are predicted for each ceRNA target, and $Y \rightarrow X$ doesn't imply $X \rightarrow Y$ because X might not be in the top 50 genes of Y . MuTaME provided 136 PTEN-regulating ceRNAs (of which 135 were targeted by LINCS) and the standalone program is not downloadable. Ala et al. used MuTaME to predict DICER1-regulating ceRNAs [35], but only 4 genes that were predicted to interact with DICER1

were targeted in LINCS. Consequently, we chose to compare the three methods when predicting PTEN regulators and targets (predictions by MuTaME are bidirectional), but genome-wide comparisons were made between Hermes and cefinder only. The comparisons suggest that Hermes outperforms MuTaME and cefinder when predicting PTEN targets and regulators (Additional file 3: Figure S2), and it significantly outperforms cefinder on genome-wide tests (Additional file 4: Figure S3).

Discussion

We proposed and re-implemented Hermes, a highly-selective context-specific method for predicting ceRNA interactions [7]. When analyzing TCGA profiles of prostate [36] and breast [37] adenocarcinomas Hermes inferred nearly 500 K ceRNA interactions for each of the two tumor types.

In total, Hermes produced expression-based evidence for the regulation of over 5000 genes by ceRNA interactions. Conclusions from perturbation assays that tested hundreds of these genes as potential targets are that half of them are dysregulated by targeting their Hermes-inferred ceRNA regulators. Put together, the results suggest that thousands of genes can be dysregulated by ceRNA interactions in each of the two contexts through exogenous perturbations or through genomic alterations that target their ceRNA regulators. We note that when evaluating interactions we chose to employ the most conservative criteria and did not correct for cell-line specific low mRNA or miRNA expression or for other types of regulation. This choice was made to ensure that inferred interactions are tested on unaltered data that was not filtered for any learned parameters. The result avoids bias and overfitting but also underestimates the predictive ability of inferred ceRNA interactions.

Our report focuses on interactions between mRNAs, where two mRNAs compete for miRNA regulation. These types of interactions have been the focus of a heated debate, with opponents suggesting that mRNA and miRNA abundances are often incompatible with physiologically relevant regulation [20]. In contrast, proponents of these interactions have described mathematical models to demonstrate their relevance [10, 38–40], in vitro experiments to support such models [41–43], as well in vivo observations of ceRNA regulation that are supported by specific in vitro validation [10, 11, 16, 17, 44].

Unlike these other reports, we present high-throughput validation of ceRNA interactions; we tested the regulation of hundreds of genes by thousands of ceRNA interaction in each of two contexts. We describe computational evidence in conjunction with high-throughput biochemical assays suggest that

ceRNA regulation is the norm and not an exception in cancer cells. While ceRNA interactions can be easily detected and validated in extreme cases—as in MYCN-amplified neuroblastomas [16] or binding-site rich RNAs [15], our results suggest that they affect the expression of thousands of genes and have the potential to synergistically dysregulate drivers of tumorigenesis in multiple tumor contexts.

Emerging evidence suggests that indirect regulation between co-regulated RNAs and even between co-regulated DNA regions is a common feature in the cell. These types of regulatory interactions can help propagate and amplify the effects of genomic alterations to dysregulate genes *in trans*, but they remain poorly understood and often overlooked. Focusing specifically on miRNA targeting, multiple groups have reported ceRNA regulation through miRNA mediators by pseudogenes [9, 44], long noncoding RNAs [12, 45] and circular RNAs [15, 40, 46, 47]. There is also growing evidence for target-mediated titration of RNA binding proteins [48–50] and transcription factors [51, 52], and most recently, evidence that changes in copy number of some chromosomal regions can alter the regulation of other regions through competition for transcription factors [53].

Conclusions

Our results suggest that ceRNA interactions—defined as indirect regulation between two mRNAs that are co-regulated by miRNAs—are significantly predictive of secondary effects observed following shRNA transfections. We report on the largest-scale effort to test these types of interactions, and our results suggest that ceRNA interactions have genome-wide effects on gene expression. Evidence from multiple studies and analysis of healthy-cell profiles suggest that ceRNA regulation is a feature of most cell types, and should not be overlooked when studying gene regulation.

Moreover, while our scope was limited, we argue that ceRNA-like regulation, including regulators, targets, and modulators that can be non-coding RNAs, proteins, and even DNA regions, can affect gene expression. These types of interactions may have weak effects on average, but these effects can alter many genes, can have strong effects in specific but possibly cases [16], and can be combinatorially amplified through multiple regulators that act as an ensemble [7, 17]. In many cases, we have limited knowledge about the biochemistry and kinetics of these interactions, but our lack of mechanistic understanding should not rule out the multitude of genetic and computational evidence for this type of regulation.

Additional files

Additional file 1: Tables S1-S5. Inferred ceRNA networks. Predicted interactions in the two tumor contexts: **Table S1.** for BRCA, and **Table S2.** for PRAD. Each table describes the coupled ceRNA pair, confidence level in the interaction, number of mediating miRNAs, and the identity of miRNA mediators. **Tables S3-4.** MCF7 and PC3 LINC5 fold changes, describing gene-expression responses to perturbagens. For each LINC5 gene profiled at a given time point after each perturbation, we provide the number of replicates (Luminex plates), observed fold change relative to control, and the standard error across plates. **Table S5.** BRCA and PRAD ceRNA validation using LINC5 data. For each LINC5 gene profiled at a given time point, we list (1) the number of predicted regulators that were silenced, (2) the size of the control set, which includes the gene's profiling after shRNA-mediated silencing of genes that were not predicted to target it, (3) observed and expected *U*-statistics, (4) and *Z*-statistics and associated the *p*-values obtained from them. For each interaction at a given time point, we list the target fold change in response to the perturbation, and the percentile rank when comparing target fold changes to fold changes of all other profiled genes in response to the perturbation. We also provide the identity of shRNA hairpins used. (PDF 1 kb)

Additional file 2: Figure S1. Accounting for systematic biases. To ensure that comparisons of target responses to predicted regulators and non-regulators are free of bias, we repeated the analysis presented in Fig. 4 after controlling for several properties. The tables demonstrate that all results, before and after controlling for the following variables were in agreement: (1) ceRNA interactions associated with the most significant ceRNA-validating responses; (2) shRNAs that can act as human miRNAs and produce off-target effects; (3) 3' UTR length; (4) 3' UTR CG content; (5) RMA-normalized expression; (6) expression variability; and (7) expression correlation with the predicted target. Here we present comparisons to results presented in Fig. 1. The number of predicted targets that were significantly ($p < 0.05$) downregulated in response to transfections of shRNAs designed to target regulators are in red; downregulated ($p > 0.05$) in orange; upregulated ($p > 0.05$) in blue, and significantly upregulated ($p < 0.05$) in green. *P* values give the confidence that the resulting distribution is not due to chance. (PDF 1027 kb)

Additional file 3: Figure S2. The effect of predicted PTEN ceRNA regulators by each of the three methods. Average PTEN mRNA fold change following shRNA-mediated silencing of its predicted regulators, as predicted by each ceRNA inference method including random assay selection, and inferences by Hermes, MuTaME, and cefinder. *P* values were calculated by comparing fold changes to random assay selection with PTEN expression profiling, using the Student's *T*-test (two-tailed). Average fold changes were normalized to the random assay selection. Bars show standard errors; * stands for $p < 0.05$; ** for $p < 0.01$; *** for $p < 0.001$. (PDF 922 kb)

Additional file 4: Figure S3. Genome wide comparison. FC comparison between Hermes, cefinder and random assay selection. Both Hermes and cefinder significantly outperform Random. Hermes outperforms cefinder at $P < 5E-07$ and $P < 2E-44$, for MCF7 and PC3, respectively; *p*-values based on two-sample Kolmogorov–Smirnov tests of ceRNA-target fold changes. (PDF 1033 kb)

Abbreviations

BRCA: Breast Invasive Carcinoma; ceRNA: competitive endogenous RNA; ceRNA: ceRNA network; LINC5: Library of integrated network-based cellular signatures; miRNA: MicroRNA; PRAD: Prostate adenocarcinomas; SEM: Standard error of the mean; TCGA: The Cancer Genome Atlas

Acknowledgements

The results published here are in part based upon data generated by The Cancer Genome Atlas pilot project established by the NCI and NHGRI as of August 2014. Information about TCGA and the investigators and institutions who constitute the TCGA research network can be found at <http://cancergenome.nih.gov/>. The dbGaP accession number for the data analyzed in this work is phs000178.

Funding

We acknowledge the generous funding provided by the NIH under the following grant awards: (1) The Outstanding NCI Investigator Award to AC (1R35CA197745), (2) the Centers for Cancer Systems Biology Consortium (1U54CA209997), (3) the instrumentation grant to AC supporting high-performance computing (1S10OD012351) and (4) the instrumentation grant to AC supporting biological data storage (1S10OD021764).

Availability of data and materials

Computational methods and analyses results are fully described here. Hermes2.0 can be downloaded free of charge from its Columbia University site <http://califano.c2b2.columbia.edu/hermes>. LINC5 assays are included in supplementary tables and can be downloaded from the LINC5 website at <http://www.linc5project.org/LINC5/data>.

Authors' contributions

HSC, PS and AC designed the study. AS and TRG designed and performed assays and contributed to their analysis. HSC, MRM, MB, XY and PS analyzed data. HSC, PS and AC wrote the manuscript. All authors read and approved the manuscript.

Competing interests

The authors declare that they have no competing interests.

Consent for publication

Not applicable.

Ethics approval and consent to participate

Not applicable.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details

¹Texas Children's Cancer Center and Department of Pediatrics, Baylor College of Medicine, Houston, TX, USA. ²IBM Research—Zurich, 8803 Rüschlikon, Zurich, Switzerland. ³Columbia Department of Systems Biology, Center for Computational Biology and Bioinformatics, Herbert Irving Comprehensive Cancer Center, Columbia University, New York, NY 10032USA. ⁴Broad Institute, 7 Cambridge Center, Cambridge, MA 02142, USA. ⁵Dana-Farber Cancer Institute, Boston, MA 02115, USA. ⁶Howard Hughes Medical Institute, Chevy Chase, MD 20815-6789, USA. ⁷MOE Key Laboratory of Bioinformatics, Tsinghua-Peking Center for Life Sciences, School of Life Sciences, Tsinghua University, Beijing 100084, China. ⁸Department of Biomedical Informatics, and Department of Biochemistry and Molecular Biophysics, and Institute for Cancer Genetics, Columbia University, New York, USA. ⁹Herbert Irving Comprehensive Cancer Center, Columbia University, New York, NY 10032, USA.

Received: 24 January 2017 Accepted: 12 May 2017

Published online: 30 May 2017

References

- Cowin PA, Anglesio M, Etemadmoghadam D, Bowtell DD. Profiling the cancer genome. *Annu Rev Genomics Hum Genet.* 2010;11:133–59.
- Califano A, Butte AJ, Friend S, Ideker T, Schadt E. Leveraging models of cell regulation and GWAS data in integrative network-based association studies. *Nat Genet.* 2012;44:841–7.
- Croce CM. Causes and consequences of microRNA dysregulation in cancer. *Nat Rev Genet.* 2009;10:704–14.
- Sumazin P, Zorman B, Chen Y, Toni-Ann M, Sarabia S, Patel K, Pollock BH, Comenford S, Finegold MJ, Tomlinson G, et al. Genomic analysis of hepatoblastoma identifies distinct molecular and prognostic subgroups. *Hepatology.* 2017;65:104–21.
- Garzon R, Calin GA, Croce CM. MicroRNAs in Cancer. *Annu Rev Med.* 2009; 60:167–79.
- Franco-Zorrilla JM, Valli A, Todesco M, Mateos I, Puga MI, Rubio-Somoza I, Leyva A, Weigel D, Garcia JA, Paz-Ares J. Target mimicry provides a new mechanism for regulation of microRNA activity. *Nat Genet.* 2007;39:1033–7.

7. Sumazin P, Yang X, Chiu HS, Chung WJ, Iyer A, Llobet-Navas D, Rajbhandari P, Bansal M, Guarnieri P, Silva J, Califano A. An extensive microRNA-mediated network of RNA-RNA interactions regulates established oncogenic pathways in glioblastoma. *Cell*. 2011;147:370–81.
8. Lee DY, Jeyapalan Z, Fang L, Yang J, Zhang Y, Yee AY, Li M, Du WW, Shatseva T, Yang BB. Expression of versican 3'-untranslated region modulates endogenous microRNA functions. *PLoS ONE*. 2010;5:e13599.
9. Poliseni L, Salmena L, Zhang J, Carver B, Haveman WJ, Pandolfi PP. A coding-independent function of gene and pseudogene mRNAs regulates tumour biology. *Nature*. 2010;465:1033–8.
10. Cazalla D, Yario T, Steitz JA. Down-regulation of a host microRNA by a Herpesvirus saimiri noncoding RNA. *Science*. 2010;328:1563–6.
11. Tay Y, Kats L, Salmena L, Weiss D, Tan SM, Ala U, Karreth F, Poliseni L, Provero P, Di Cunto F, et al. Coding-independent regulation of the tumor suppressor PTEN by competing endogenous mRNAs. *Cell*. 2011;147:344–57.
12. Cesana M, Cacchiarelli D, Legnini I, Santini T, Sthandier O, Chinappi M, Tramontano A, Bozzoni I. A long noncoding RNA controls muscle differentiation by functioning as a competing endogenous RNA. *Cell*. 2011;147:358–69.
13. Tay Y, Rinn J, Pandolfi PP. The multilayered complexity of ceRNA crosstalk and competition. *Nature*. 2014;505:344–52.
14. Kumar MS, Armenteros-Monterroso E, East P, Chakravorty P, Matthews N, Winslow MM, Downward J. HMG2A functions as a competing endogenous RNA to promote lung cancer progression. *Nature*. 2014;505:212–7.
15. Hansen TB, Jensen TI, Clausen BH, Bramsen JB, Finsen B, Damgaard CK, Kjems J. Natural RNA circles function as efficient microRNA sponges. *Nature*. 2013;495:384–8.
16. Powers JT, Tzanov KM, Pearson DS, Roels F, Spina CS, Ebright R, Seligson M, de Soysa Y, Cahan P, Theissen J. Multiple mechanisms disrupt the let-7 microRNA family in neuroblastoma. *Nature*. 2016;535:246–51.
17. Chiu HS, Llobet-Navas D, Yang X, Chung WJ, Ambesi-Impiomato A, Iyer A, Kim HR, Seviour EG, Luo Z, Sehgal V, et al. Cupid: simultaneous reconstruction of microRNA-target and ceRNA networks. *Genome Res*. 2015; 25:257–67.
18. Denzler R, Agarwal V, Stefano J, Bartel David P, Stoffel M. Assessing the ceRNA Hypothesis with Quantitative Measurements of miRNA and Target Abundance. *Mol Cell*. 2014;54:766–76.
19. Wee LM, Flores-Jasso CF, Salomon WE, Zamore PD. Argonaute divides its RNA guide into domains with distinct functions and RNA-binding properties. *Cell*. 2012;151:1055–67.
20. Denzler R, McGeary SE, Title AC, Agarwal V, Bartel DP, Stoffel M. Impact of MicroRNA Levels, Target-Site Complementarity, and Cooperativity on Competing Endogenous RNA-Regulated Gene Expression. *Molecular Cell*. 2016;64:565–79.
21. Peck D, Crawford ED, Ross KN, Stegmaier K, Golub TR, Lamb J. A method for high-throughput gene expression signature analysis. *Genome Biol*. 2006;7:R61.
22. Golub TR. The Somatic Cancer Genome. *Blood*. 2014;124:SC14–4.
23. Lamb J, Golub TR, Subramanian A, Peck DD. Gene-expression profiling with reduced numbers of transcript measurements. In: Google Patents; 2013.
24. Sarver AL, Subramanian S. Competing endogenous RNA database. *Bioinformatics*. 2012;8:731–3.
25. Darbellay G, Vajda I. Estimation of the Information by an Adaptive Partitioning of the Observation Space. *IEEE Trans on Information Theory*. 1999;45:1315–21.
26. Buehler E, Khan AA, Marine S, Rajaram M, Bahl A, Burchard J, Ferrer M. siRNA off-target effects in genome-wide screens identify signaling pathway members. *Sci Rep*. 2012;2:428.
27. Petryszak R, Keays M, Tang YA, Fonseca NA, Barrera E, Burdett T, Füllgrabe A, Fuentes AM-P, Jupp S, Koskinen S. Expression Atlas update—an integrated database of gene and protein expression in humans, animals and plants. *Nucleic Acids Res*. 2016;44(D1):D746–752.
28. Marsaglia G, Tsang W, Wang J. Evaluating Kolmogorov's Distribution. *J Stat Softw*. 2003;8:1–4.
29. Taylor BS, Schultz N, Hieronymus H, Gopalan A, Xiao Y, Carver BS, Arora VK, Kaushik P, Cerami E, Reva B, et al. Integrative genomic profiling of human prostate cancer. *Cancer Cell*. 2010;18:11–22.
30. Buffa FM, Camps C, Winchester L, Snell CE, Gee HE, Sheldon H, Taylor M, Harris AL, Ragoussis J. microRNA-associated progression pathways and potential therapeutic targets identified by integrated mRNA and microRNA expression profiling in breast cancer. *Cancer Res*. 2011;71:5635–45.
31. Basso K, Margolin AA, Stolovitzky G, Klein U, Dalla-Favera R, Califano A. Reverse engineering of regulatory networks in human B cells. *Nat Genet*. 2005;37:382–90.
32. Wang K, Saito M, Bisikirska BC, Alvarez MJ, Lim WK, Rajbhandari P, Shen Q, Nemenman I, Basso K, Margolin AA, et al. Genome-wide identification of post-translational modulators of transcription factor activity in human B cells. *Nat Biotechnol*. 2009;27:829–39.
33. Lefebvre C, Rajbhandari P, Alvarez MJ, Bandar P, Lim WK, Sato M, Wang K, Sumazin P, Kustagi M, Bisikirska BC, et al. A human B-cell interactome identifies MYB and FOXM1 as master regulators of proliferation in germinal centers. *Mol Syst Biol*. 2010;6:377.
34. Carro MS, Lim WK, Alvarez MJ, Bollo RJ, Zhao X, Snyder EY, Sulman EP, Anne SL, Doetsch F, Colman H, et al. The transcriptional network for mesenchymal transformation of brain tumours. *Nature*. 2010;463:318–25.
35. Ala U, Karreth FA, Bosia C, Pagnani A, Taulli R, Léopold V, Tay Y, Provero P, Zecchina R, Pandolfi PP. Integrated transcriptional and competitive endogenous RNA networks are cross-regulated in permissive molecular environments. *Proc Natl Acad Sci*. 2013;110:7154–9.
36. The Cancer Genome Atlas. The molecular taxonomy of primary prostate cancer. *Cell*. 2015;163:1011–25.
37. The Cancer Genome Atlas. Comprehensive molecular portraits of human breast tumours. *Nature*. 2012;490:61–70.
38. Arvey A, Larsson E, Sander C, Leslie CS, Marks DS. Target mRNA abundance dilutes microRNA and siRNA activity. *Mol Syst Biol*. 2010;6:363.
39. Bosson AD, Zamudio JR, Sharp PA. Endogenous miRNA and target concentrations determine susceptibility to potential ceRNA competition. *Mol Cell*. 2014;56:347–59.
40. Memczak S, Jens M, Elefimioti A, Torti F, Krueger J, Rybak A, Maier L, Mackowiak SD, Gregersen LH, Munschauer M. Circular RNAs are a large class of animal RNAs with regulatory potency. *Nature*. 2013;495:333–8.
41. Bosson Andrew D, Zamudio Jesse R, Sharp Phillip A. Endogenous miRNA and Target Concentrations Determine Susceptibility to Potential ceRNA Competition. *Mol Cell*. 2014;56(3):347–59.
42. Luna JM, Scheel TK, Danino T, Shaw KS, Mele A, Fak JJ, Nishiuchi E, Takacs CN, Catanese MT, de Jong YP, et al. Hepatitis C Virus RNA Functionally Sequesters miR-122. *Cell*. 2015;160:1099–110.
43. Yuan Y, Liu B, Xie P, Zhang MQ, Li Y, Xie Z, Wang X. Model-guided quantitative analysis of microRNA-mediated regulation on competing endogenous RNAs using a synthetic gene circuit. *Proc Natl Acad Sci U S A*. 2015;112:3158–63.
44. Karreth FA, Reschke M, Ruocco A, Ng C, Chapuy B, Léopold V, Sjöberg M, Keane TM, Verma A, Ala U. The BRAF pseudogene functions as a competitive endogenous RNA and induces lymphoma in vivo. *Cell*. 2015; 161:319–32.
45. Wang J, Liu X, Wu H, Ni P, Gu Z, Qiao Y, Chen N, Sun F, Fan Q. CREB up-regulates long non-coding RNA, HULC expression through interaction with microRNA-372 in liver cancer. *Nucleic Acids Res*. 2010;38:5366–83.
46. Li J-H, Liu S, Zhou H, Qu L-H, Yang J-H. starBase v2.0: decoding miRNA-ceRNA, miRNA-ncRNA and protein-RNA interaction networks from large-scale CLIP-Seq data. *Nucleic Acids Res*. 2014;42(Database issue):D92–97.
47. Bhattacharya A, Cui Y. SomamiR 2.0: a database of cancer somatic mutations altering microRNA-ceRNA interactions. *Nucleic Acids Res*. 2016; 44(D1):D1005–1010.
48. Donnelly CJ, Willis DE, Xu M, Tep C, Jiang C, Yoo S, Schanen NC, Kirn-Safran CB, Van Minnen J, English A. Limited availability of ZBP1 restricts axonal mRNA localization and nerve regeneration capacity. *EMBO J*. 2011;30:4665–77.
49. Sosanya NM, Huang PP, Cacheaux LP, Chen CJ, Nguyen K, Perrone-Bizzozero NI, Raab-Graham KF. Degradation of high affinity HuD targets releases Kv1.1 mRNA from miR-129 repression by mTORC1. *J Cell Biol*. 2013;202:53–69.
50. Kim J, Abdelmohsen K, Yang X, De S, Grammatikakis I, Noh JH, Gorospe M. LncRNA OIP5-AS1/cyano sponges RNA-binding protein HuR. *Nucleic Acids Res*. 2016;44:2378–92.
51. Hung T, Wang Y, Lin MF, Koegel AK, Kotake Y, Grant GD, Horlings HM, Shah N, Umbrecht C, Wang P. Extensive and coordinated transcription of noncoding RNAs within cell-cycle promoters. *Nat Genet*. 2011;43:621–9.
52. Wang KC, Chang HY. Molecular mechanisms of long noncoding RNAs. *Mol Cell*. 2011;43:904–14.
53. Jones DL, Brewster RC, Phillips R. Promoter architecture dictates cell-to-cell variability in gene expression. *Science*. 2014;346:1533–6.