MIT Open Access Articles

Deconvolution of seed and RNA-binding protein crosstalk in RNAi-based functional genomics

Massachusetts Institute of Technology

DSpace@MIT

Deconvolution of seed and RNA-binding protein crosstalk in RNAi-based functional genomics

Hiroshi I. Suzuki[1], **Ryan M. Spengler**[2], **Giedre Grigelioniene**[3,4,5], **Tatsuya Kobayashi**[3], and **Phillip A. Sharp**[1,6,*]

[1]David H. Koch Institute for Integrative Cancer Research, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA

[2]Department of Internal Medicine, University of Michigan, Ann Arbor, Michigan, USA

[3]Endocrine Unit, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts, USA

[4]Department of Molecular Medicine and Surgery and Center for Molecular Medicine, Karolinska Institutet, Stockholm, Sweden

[5]Department of Clinical Genetics, Karolinska University Hospital, Stockholm, Sweden

[6]Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA

Abstract

RNA interference (RNAi) is a major powerful platform for gene perturbations, but is restricted by off-target mechanisms. Communication between RNAs, small RNAs, and RNA-binding proteins (RBPs) is a pervasive feature of cellular RNA networks. We present a new crosstalk scenario, designated as "crosstalk with endogenous RBPs (ceRBP)", where siRNAs or miRNAs with seed sequences that overlap RBP motifs have extended biological effects by perturbing endogenous RBP activity. Systematic analysis of siRNA off-target data and genome-wide RNAi cancer lethality screens using 501 human cancer cell lines, a cancer dependency map, revealed that seed-to-RBP crosstalk is widespread, contributes to off-target activity, and affects RNAi performance. Specifically, deconvolution of the interactions between gene knockdown and seed-mediated silencing effects in the cancer dependency map showed widespread contributions of seed-to-RBP

*Corresponding author: Phillip A. Sharp, Koch Institute for Integrative Cancer Research, Massachusetts Institute of Technology, 500 Main St., 76-461A, Cambridge, MA 02139; Phone: 617-253-6421; FAX: 617-253-3867; sharppa@mit.edu.

crosstalk to growth-phenotype modulation. These findings suggest a novel aspect of miRNA biology and offer a basis for improvement of RNAi agents and RNAi-based functional genomics.

RNA interference (RNAi) is a major powerful platform for large-scale gene perturbations, but is limited by off-target mechanisms. Since RNAi utilizes the endogenous microRNA (miRNA) pathway that depends on the seed sequence for target RNA recognition, RNAi-based experimental perturbation can be confounded by miRNA seed-based off-target effects. Several computational methods have been developed to enrich on-target verses off-target effects[1-4]. Crosstalk between mRNAs, small RNAs, and RNA-binding proteins (RBPs) is a pervasive feature of cells. Recent advances in miRNA biology have identified several crosstalk scenarios: (1) ceRNAs (competing endogenous RNAs) targeted by the same miRNA can show indirect positively correlated expression by titration of miRNA activities[5]; and (2) multiple RBPs can modulate miRNA activities targeting the same RNAs[6]. However, the extent of seed-based crosstalk with RBPs has not been analyzed in interpretation of RNAi-based functional genomics.

We recently discovered a crosstalk scenario in which a novel seed generated by mutation of a miRNA gene overlaps and perturbs the activity of an RBP (G.G. and H.I.S. et al., unpublished data). Numerous evidences have established that conserved functions of endogenous miRNAs are a product of co-evolution of miRNA and target networks[7]. We found that mutation of a super-enhancer-associated conserved miRNA causes a human congenital skeletal disorder and that the mutant miRNA in mice elicits potent target repression through a new repertoire of poorly conserved seed targets. The robustness of the mutant phenotype can be explained by the mutant miRNA competing with the conserved RNA-binding protein YBX1 and perturbing its activities. These observations suggest that small RNAs with seeds overlapping the RBP binding motifs can have extended biological roles, i.e. gain of function, by interfering with RBP activities (Fig. 1a).

We hypothesized that this new seed-RBP crosstalk mechanism ("crosstalk with endogenous RBPs (ceRBP)") may be prevalent in RNAi and important for interpretation of RNAi off-target phenotypes. To this end, we performed systematic analysis of two large datasets: (1) 93 published microarray datasets that monitored the off-target response of transfected siRNAs (Fig. 1b-d)[8]; and (2) large-scale RNAi cell lethality screens using ~100,000 shRNAs and 501 cancer cell lines (Fig. 2-4)[3].

Analysis of the relationship between off-target effects and seed-RBP motif overlap using the microarray results for 93 siRNAs[8,9] revealed that siRNAs with seeds overlapping these motifs (218 RBP binding motifs defined by CIS-BP-RNA database, including the results of large-scale experiments for *in vitro* sequence preferences of RBPs[10]) had higher numbers of downregulated off-target genes than those that did not overlap, while upregulation of off-targets was a minor change (Fig. 1b and Supplementary Fig. 1). RBP motif-overlapping siRNAs were also associated with higher changes in background genes in both down- and upregulation directions, suggesting higher secondary transcriptome effects (Fig. 1c). In addition, RBP motif-overlapping siRNAs were associated with high values of the previously defined potential off-targeting score (POTS)[8] (Fig. 1d). Conversely, a large fraction of

siRNAs with low off-target effects avoided RBP motif overlap, suggesting this minimizes RNAi off-target effects.

Next, we processed a recently reported database of Cancer Dependency Map by Tsherniak et al.[3], which integratively analyzed genome-scale RNAi screens. This project quantitated the effect on growth of ~100,000 shRNAs targeting ~17,000 genes in 501 diverse human cancer cell lines. After 16 population doublings or 40 days, genes targeted by the depleted shRNAs are inferred to be essential for cell viability or proliferation. They reported that shRNA depletion scores for shRNA pairs with the same seeds were more correlated than those for shRNA pairs targeting the same gene, indicating that seed effects are highly prevalent in such screens[3]. The new computational model, DEMETER, was developed to segregate direct effects of suppressing the intended genes from off-target effects caused by expression of seed sequences (Fig. 2a)[3]. We used this DEMETER model to validate the ceRBP concept. DEMETER provides gene-level dependency scores across 501 cell lines for ~17,000 genes and seed-level dependency scores for ~15,000 seeds[3]. We computed the correlation coefficient between gene-level and seed-level dependency scores to examine how seed-induced phenotypes correlate with RBP knockdown-induced phenotypes (Fig. 2b). When small RNAs and RBPs with shared binding sites functionally compete, a positive correlation between RBP knockdown phenotypes and seed overexpression phenotypes would be observed (Fig. 2b). On the other hand, when they cooperate and/or are redundant, a negative correlation between two dependency scores, i.e. phenocopy phenomenon, would be observed.

We analyzed motif overlap with 166 human RBPs in CIS-BP-RNA database[10] and found that about 70 RBPs show a significant competitive or cooperative relationship with motif-overlapping seeds (Fig.2c and Supplementary Table 1), suggesting that ceRBP effects from shRNAs are widespread. Consistent with our recent findings (G.G. and H.I.S. et al., unpublished data), the competitive relationship between the RBP YBX1 and overlapping seeds was confirmed (Fig. 2d). Competitive relationships with shRNA seeds were indicated for several important RBPs including IGF2BP1/2/3, ELAVL1(HuR)/ELAV4(HuD), and FXR1 (Fig. 2d and Supplementary Fig. 2). Competition with known splicing regulators RBM28, PTBP1, RBFOX2, and ESRP2 was also observed, but these factors shuttle between the nucleus and cytoplasm. Coherently, some RBP families such as IGF2BP1/2/3 and ELAVL1/4 showed a similar competitive relationship with multiple seeds. IGF2BPs, ELAVL1, and YBX1 stabilize target RNAs and it has been reported that IGF2BPs and ELAVL1 antagonize miRNA regulation on multiple miRNA targets[11-14]. Conversely, for several RBPs including PABPC1, RBM47, and EIF4B, cooperative and/or redundant relationships with overlapping seeds were observed (Fig. 2d and Supplementary Fig. 2). Poly(A)-binding protein PABP has been reported to stimulate miRNA-mediated repression[15]. This may be probably a consequence of overlapping binding sites. Overall, these reports are largely consistent with our model, supporting the validity of the analysis. Interestingly, a surprising number of these RBPs, including IGF2BPs, ELAVL1, PTBP1, RBM28, RBM47, RBFOX2, and ESRP2, are abnormally regulated in cancer[16].

We performed experimental validation of interactions between IGF2BPs and RBP motif-overlapping seeds. Based on previous PAR-CLIP and RNA expression datasets[17,18], we

focused on a representative IGF2BP target gene *HMGA2* (Supplementary Fig. 3). RNA expression and 3′ UTR reporter analysis confirmed that shRNAs with IGF2BP2 motif-overlapping seeds target *HMGA2* mRNA and 3′ UTR (Fig. 3a and 3b). Furthermore, a single cell reporter system using *Igf2bp1/2/3* wild-type and triple KO mesenchymal stem cells (*Dicer* KO and human Dicer (hsDicer)-rescued background) recapitulated the results of 3′ UTR luciferase reporter analysis and showed that suppressive effects of RBP motif-overlapping seeds are higher in wild-type cells relative to triple KO cells (Fig. 3c, 3d and Supplementary Fig. 4)[18,19]. This suggests that susceptibility of *HMGA2* 3′ UTR to overlapping seeds increases in the presence of IGF2BPs, reinforcing our concept.

We next analyzed roles of the ceRBP model in shRNA performance in the Cancer Dependency Map database. As stated above, when we compare the contribution of gene effect and seed effect for each of ~100,000 shRNAs, seed effects were more prevalent than gene effects for cell viability or proliferation (Fig. 4a, left). In Fig. 4a, shRNAs with high on-target specificity and low seed off-target effects would be primarily located in the top-left region, and shRNAs with high seed off-target effects would be enriched in the bottom-right region. For several competitive RBPs including IGF2BP2, RBM28, ESRP2, and RBFOX2, shRNAs with seeds overlapping their binding motifs seldom have high gene effects and low seed effects, i.e. features of good shRNAs (Fig. 4a, middle, and Supplementary Fig. 5 and 6). In contrast, for several cooperative RBPs including PABPC1 and RBM47, shRNAs with overlapping seeds can have high gene effects and low seed effects (Fig. 4a, right, and Supplementary Fig. 5). This apparent high efficiency is complicated as they might perturb the RNP regulatory network of these proteins. It is also the case that avoiding seeds overlapping competitive RBPs will enrich shRNAs with low seed effects and high gene effects. To better illustrate the above, we next quantitatively estimated the impact of competitive or cooperative overlapping seeds for a subset of shRNAs. We focused on shRNAs against selected cancer essential genes differentially required for growth of three or more cell lines[3]. Analysis of 2,690 shRNAs targeting 448 cancer genes revealed that competitive seeds clearly decrease the frequency of very efficient shRNAs about 1.5-fold (the top-left region, 27.8% to 18.5 %), while cooperative seeds show no decrease (Fig. 4b and Supplementary Fig. 7). These findings indicate that competition with RBPs substantially decreases the efficiency and specificity of RNAi.

Finally, we analyzed the specificity of shRNAs with seeds identical to endogenous miRNAs and found that overlap with the broadly conserved miRNAs results in an approximately 2-fold decrease in gene effects in a q1 group, i.e. a low seed effect population (Supplementary Fig. 8)[20]. In addition, seeds of approximately 25 % of all broadly conserved miRNAs overlap RBP motifs (Supplementary Fig. 9 and Supplementary Table 2). The conserved and poorly conserved miRNAs show slight higher overlap approximately 30%. This compares with 20% overlap with all ~15,000 seeds considered in the above data.

In summary, this proof-of-concept analysis suggests that seed-to-RBP crosstalk is prevalent and contributes to seed-induced off-target effects and growth phenotype modulation. The results have three implications for RNAi design, miRNA biology, and the interpretation of functional genomics. First, avoiding shRNAs with seeds that overlap with RBP binding motifs will enhance development of RNAi agents that have effective silencing of the target

mRNA with less seed-mediated off-target effects. This will probably translate to better experimental and therapeutic agents. Second, ceRBP concepts add another layer of regulation to miRNA biology. We observed that a fraction of endogenous miRNA seeds potentially crosstalk with endogenous RBPs. For example, this includes a probable competition between oncogenic IGF2BP2 and tumor suppressive miR-147 (Supplementary Table 2)[21]. IGF2BPs are also thought to antagonize let-7 miRNAs[11,12], but this is apparently not due to seed overlap. A major limitation in this analysis of overlap of RBP sites and shRNA seeds is the poorly defined state of the former. The RBP motifs are statistically validated but not all motif sites in RNA bind the protein and many RBPs bind at sites with only partial motifs. Thus, this analysis is certainly correct when viewed as a population but also certainly has limitations when considered for particular RNA sites. Further updating the features of RBP interactions will increase the predictability of this type of regulation. In addition, seed-RBP interactions in the Cancer Dependency Map database analysis are based on cellular phenotype evaluation, and multiple modes of mechanistic crosstalk between AGO and RBPs can result in phenotypic resemblance. Third, the DEMETER model and our analysis suggest that an advance in data deconvolution processes can not only improve gene dependency profiling, but also simultaneously provide data resources of gene knockdown and seed effects, and their relationships. Seed and gene correlation analysis can be extended to analysis using all genes not just RBPs, possibly leading to detection of unknown seed-gene relationships and biologically relevant miRNA-target relationships.

Methods

Methods, including statements of data availability and any associated accession codes and references, are available in the online version of the paper.

Online Methods

RBP binding motif analysis

To detect the overlap between RNA sequences and RBP binding motifs, we performed RNA binding protein motif analysis using the CIS-BP-RNA database (see URLs section)[10]. CIS-BP-RNA database, the Catalog of Inferred Sequence Binding Proteins of RNA, is a library of RNA binding protein (RBP) motifs and specificities[10]. Human RBP datasets including 166 RBPs and 218 unique motifs were used, and the overlap was determined with "PWMs – Log Odds" scoring system with a default threshold.

siRNA off-target analysis

For siRNA off-target analysis, we examined a relationship between siRNA-RBP crosstalk and degrees of siRNA off-target effects using 93 siRNA transfection microarray experiments described in a siSPOTR development report[8]. Pre-processed microarray datasets were obtained from a previous report[9]. For all siRNA sequences, RNA binding protein motif analysis was performed using the CIS-BP-RNA database[10], and overlap between siRNA complementary sequences (seed position: 2-8) and RBP motifs more than two bases was considered as a possible indicator of RBP crosstalk. In Fig. 1b, for each siRNA, numbers of downregulated or upregulated off-targets (transcripts with 3′ UTRs containing 7- and 8-mer

seed-binding sites and −0.3 or 0.3 log2 fold change) over background were analyzed. In Fig. 1c, proportion of downregulated and upregulated background genes with no predicted 3′ UTR seed-binding sites for either siRNA strand was analyzed. According to siSPOTR siRNA design tool[8], potential off-target score (POTS) values were also analyzed (Fig. 1d). Low POTS values (especially < 50) indicate low off-targeting potential. In Fig. 1b, 1d, and Supplementary Fig. 1, statistical significance was assessed using two-sided Wilcoxon signed rank test. In Fig. 1c, statistical significance was assessed using one-sided Welch's t test.

DEMETER-ceRBP analysis

DEMETER gene knockdown effect for 17098 genes (dataset 1), DEMETER seed effects for 15142 7mer sequences (dataset 2), and DEMETER shRNA performance metrics for ~100,000 shRNAs (dataset 3) are from Project Achilles database (see URLs section)[3]. For DEMETER-ceRBP analysis, Pearson correlations of dataset 1 and dataset 2 were computed across 501 cell lines for all gene-seed pairs. For all complementary sequences of 7mer sequences of shRNA guide strands (positions 12-18 (Seed2) and 11-17 (Seed1) on the sense strand), RNA binding protein motif analysis was performed using the CIS-BP-RNA database[10], and overlap between 7mer sequences and entire regions of RBP motifs was considered. For each RBP, Pearson correlation coefficient of pairs between indicated RBP and all seeds were converted to Z-scores. Then, the median changes between all seeds and motif-overlapping seeds were calculated for each RBP (Fig. 2c and Supplementary Table 1). P values were calculated by one-sided Kolmogorov–Smirnov (K-S) test for either direction depending on the median changes (Fig. 2c,d, Supplementary Fig. 2, and Supplementary Table 1).

DEMETER shRNA performance analysis

We analyzed the variance scores explained by the contribution of the gene effect and seed effect (gene solution and seed solutions considering both Seed1 and Seed2, respectively) in DEMETER shRNA performance metrics for ~100,000 shRNAs (dataset 3) (Fig. 4a and Supplementary Fig. 6 and 8)[3]. In Fig. 4b, among 769 gene dependencies for which the DEMETER scores of at least one cell line were six SDs or greater from the mean, we analyzed shRNAs for 448 cancer genes for which the DEMETER scores of at least three cell lines were above six SDs to increase the reliability of quantitative assessment of shRNA performance. shRNAs that both Seed1 and Seed2 did not overlap with any RBP motifs were selected as shRNAs with RBP motif non-overlapping seeds. For shRNAs with RBP motif-overlapping seeds, overlap with either Seed1 or Seed2 was considered. Competitive seeds are seeds overlapping the motifs of statistically significant RBPs in DEMETER-ceRBP analysis (competition direction in Fig. 2c, P < 0.05) and having Z-score > 1 (Fig. 4b) or 2 (Supplementary Fig. 7). Cooperative seeds are seeds overlapping the motifs of statistically significant RBPs in DEMETER-ceRBP analysis (cooperation direction in Fig. 2c, P < 0.05) and having Z-score < -1 (Fig. 4b) or -2 (Supplementary Fig. 7). In Supplementary Fig. 6 and 8, shRNAs were grouped into three groups according to values of DEMETER seed solutions (q1: < 0.2, q2: 0.2 and < 0.6, and q3: 0.6). In Supplementary Fig. 6, analysis of all shRNAs showed that shRNAs with competitive seeds have significantly lower gene effects in a group with low seed effects (q1 group). In Supplementary Fig. 8, shRNAs with seeds overlapping seeds of broadly conserved miRNAs show an approximately 2-fold decrease in

gene effects in a q1 group. In Supplementary Fig. 6 and 8, statistical significance was assessed using two-sided Wilcoxon signed rank test with Bonferroni correction.

miRNA information

Information of human miRNAs were downloaded from the TargetScan database (v7.1, see URLs section)[20] in August of 2017. Classification of miRNAs (broadly conserved (2), conserved (1), poorly conserved (0), and others (-1)) is according to TargetScan information. Overlap between shRNA 7mer and miRNA seed+m8 sequences was analyzed.

IGF2BP target analysis and shRNA design

IGF2BP1/2/3 PAR-CLIP datasets, IGF2BP1/2/3 triple knockdown microarray datatasets, and *Igf2bp1/2/3* triple knockout RNAseq datasets were previously described[17,18]. In the report by Hafner et al.[17], IGF2BP1/2/3 knockdown microarray experiments in HEK293 cells have shown that IGF2BPs tend to globally stabilize the target transcripts. We confirmed this observation (Supplementary Fig. 3, left) and further interrogated whether these transcripts are similarly regulated in mouse mesenchymal stem cells (MSCs) using the datasets from JnBaptiste et al.[18]. Re-analysis of RNAseq datasets in *Igf2bp1/2/3* wild-type and triple knockout MSCs revealed that IGF2BP1/2/3 target transcripts are similarly destabilized by triple knockout of IGF2BPs in MSCs (Supplementary Fig. 3, middle), suggesting conservation of these targets. We focused on the representative target *HMGA2* and selected shRNAs with IGF2BP2 motif-overlapping seeds and RBP motif non-overlapping seeds for subsequent functional analyses as shown in Supplementary Fig. 3, the right panel. Seed2 was considered for selection of shRNAs with IGF2BP2 motif-overlapping seeds. Selected shRNAs were cloned into pGL3-U6-shRNA-PGK-blastR plasmids using AgeI and EcoRI sites. The shRNA sequences are described in Supplementary Table 3. In Supplementary Fig. 3, P values were calculated by two-sided Wilcoxon signed rank test with Bonferroni correction.

qRT-PCR analysis

HEK293T cells were from American Type Culture Collection and maintained in a Dulbecco's Modified Eagle's Medium (Thermo Fisher Scientific) supplemented with 10% fetal bovine serum (FBS), penicillin, and streptomycin. HEK293T cells were transfected with pGL3-U6-shRNA-PGK-blastR shRNA expression plasmids and selected with Blasticidin S (15 μg/ml) for 3 days. After selection, total RNAs were extracted using TRIzol Reagent (Thermo Fisher Scientific) and analyzed by qRT-PCR analysis using SuperScript III First-Strand Synthesis System, PowerUp SYBR Green Master Mix (Thermo Fisher Scientific), and the 7500 Fast Real-Time PCR System (Applied Biosystems). Results were normalized to β-actin expression. Statistical significance was assessed using two-sided Wilcoxon signed rank test. Primer sequences used are described in Supplementary Table 3.

Dual luciferase reporter assay

Human *HMGA2* 3′ UTR was cloned into the 3′ UTR of the luciferase gene in psiCHECK 2 dual luciferase reporter vector (Promega) using XhoI and NotI sites. HEK293T cells were transfected with reporter and shRNA plasmids using Lipofectamine 2000 (Thermo Fisher

Scientific). 48 hr after transfection, the ratio between firefly and *Renilla* luciferase was determined using Dual-Luciferase Reporter Assay System (Promega). Statistical significance was assessed using two-sided Wilcoxon signed rank test. Primer sequences used are described in Supplementary Table 3.

Single cell dual fluorescence reporter assay

The bidirectional pTRE-Tight-BI (Clontech) eYFP and mCherry reporter vector has been previously described[19]. Human *HMGA2* 3′ UTR was cloned into the 3′ UTR of mCherry using ClaI and SalI sites. Primer sequences used are described in Supplementary Table 3. *Dicer* KO, human Dicer (hsDicer)-rescued, and *Igf2bp1/2/3* WT or triple KO (tKO) mouse mesenchymal stem cells (MSC) have been previously described[18]. For *Igf2bp1/2/3* tKO MSC, clone 5 was used. As previously described[18], *Dicer* KO and human Dicer (hsDicer)-rescued MSCs overexpress *Igf2bp1/2/3* and *Hmga2.* Upon triple knockout of *Igf2bp1/2/3*, mRNA expression levels of *Hmga2* were downregulated about 50 %, probably reflecting IGF2BP-mediated stabilization of *Hmga2* mRNA[18]. MSCs were maintained in α-MEM supplemented with 10% fetal bovine serum (FBS), penicillin, and streptomycin. MSCs were transfected with eYFP and mCherry reporter, rtTA expression, and shRNA expression plasmids using Lipofectamine 2000 (Thermo Fisher Scientific). 4 hr after transfection, cells were treated with 1 μg/ml doxycycline (Sigma). 48 hr after transfection, flow cytometry analysis was performed using BD FACS Celesta (BD Biosciences). Data was collected using FACS Diva Version 8.0.1 and analyzed using FlowJo version 10.4.1. and R. After gating for single cell populations (Supplementary Fig. 4), about 10,000 eYFP-positive cells were analyzed. Each experiment included untransfection controls. The eYFP and mCherry signals of each cell were background normalized by subtracting the mean plus two standard deviation of signal in untransfection samples and binned by eYFP signal levels as previously described[5,19].

Statistical analysis

Statistical tests were performed using JMP or R. In molecular biology experiments, results are representative of more than two independent and reproducible experiments. In Fig. 1b, 1d, and Supplementary Fig. 1, statistical significance was assessed using two-sided Wilcoxon signed rank test. In Fig. 1c, statistical significance was assessed using one-sided Welch's t test. In DEMETER-ceRBP analysis (Fig. 2c,d, Supplementary Fig. 2, and Supplementary Table 1), P values were calculated by one-sided Kolmogorov–Smirnov (K-S) test for either direction depending on the median changes. In Fig. 3a and 3b, statistical significance was assessed using two-sided Wilcoxon signed rank test. In Supplementary Fig. 3, 6, and 8, P values were calculated by two-sided Wilcoxon signed rank test with Bonferroni correction. In all box plots, center lines show medians; box limits indicate the twenty-fifth and seventy-fifth percentiles; whiskers extend to 1.5× the interquartile range.

Life Sciences Reporting Summary

Further information on experimental design is available in the Life Sciences Reporting Summary.

Data availability

The datasets analyzed in the current study are available from previous reports[8,9], Project Achilles database[3], and Cancer Dependency Map[3].

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

1. Konig R, et al. A probability-based approach for the analysis of large-scale RNAi screens. Nat Methods. 2007; 4:847–9. [PubMed: 17828270]

2. Shao DD, et al. ATARiS: computational quantification of gene suppression phenotypes from multisample RNAi screens. Genome Res. 2013; 23:665–78. [PubMed: 23269662]

3. Tsherniak A, et al. Defining a Cancer Dependency Map. Cell. 2017; 170:564–576 e16. [PubMed: 28753430]

4. McDonald ER 3rd, et al. Project DRIVE: A Compendium of Cancer Dependencies and Synthetic Lethal Relationships Uncovered by Large-Scale, Deep RNAi Screening. Cell. 2017; 170:577–592 e10. [PubMed: 28753431]

5. Bosson AD, Zamudio JR, Sharp PA. Endogenous miRNA and target concentrations determine susceptibility to potential ceRNA competition. Mol Cell. 2014; 56:347–59. [PubMed: 25449132]

6. van Kouwenhove M, Kedde M, Agami R. MicroRNA regulation by RNA-binding proteins and its implications for cancer. Nat Rev Cancer. 2011; 11:644–56. [PubMed: 21822212]

7. Berezikov E. Evolution of microRNA diversity and regulation in animals. Nat Rev Genet. 2011; 12:846–60. [PubMed: 22094948]

8. Boudreau RL, et al. siSPOTR: a tool for designing highly specific and potent siRNAs for human and mouse. Nucleic Acids Res. 2013; 41:e9. [PubMed: 22941647]

9. Garcia DM, et al. Weak seed-pairing stability and high target-site abundance decrease the proficiency of lsy-6 and other microRNAs. Nat Struct Mol Biol. 2011; 18:1139–46. [PubMed: 21909094]

10. Ray D, et al. A compendium of RNA-binding motifs for decoding gene regulation. Nature. 2013; 499:172–7. [PubMed: 23846655]

11. Jonson L, et al. IMP3 RNP safe houses prevent miRNA-directed HMGA2 mRNA decay in cancer and development. Cell Rep. 2014; 7:539–51. [PubMed: 24703842]

12. Degrauwe N, et al. The RNA Binding Protein IMP2 Preserves Glioblastoma Stem Cells by Preventing let-7 Target Gene Silencing. Cell Rep. 2016; 15:1634–47. [PubMed: 27184842]

13. Goodarzi H, et al. Endogenous tRNA-Derived Fragments Suppress Breast Cancer Progression via YBX1 Displacement. Cell. 2015; 161:790–802. [PubMed: 25957686]

14. Lu YC, et al. ELAVL1 modulates transcriptome-wide miRNA binding in murine macrophages. Cell Rep. 2014; 9:2330–43. [PubMed: 25533351]

15. Moretti F, Kaiser C, Zdanowicz-Specht A, Hentze MW. PABP and the poly(A) tail augment microRNA repression by facilitated miRISC binding. Nat Struct Mol Biol. 2012; 19:603–8. [PubMed: 22635249]

16. Sebestyen E, et al. Large-scale analysis of genome and transcriptome alterations in multiple tumors unveils novel cancer-relevant splicing networks. Genome Res. 2016; 26:732–44. [PubMed: 27197215]

17. Hafner M, et al. Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. Cell. 2010; 141:129–41. [PubMed: 20371350]

18. JnBaptiste CK, et al. Dicer loss and recovery induce an oncogenic switch driven by transcriptional activation of the oncofetal Imp1-3 family. Genes Dev. 2017; 31:674–687. [PubMed: 28446596]

19. Mukherji S, et al. MicroRNAs can generate thresholds in target gene expression. Nat Genet. 2011; 43:854–9. [PubMed: 21857679]

20. Agarwal V, Bell GW, Nam JW, Bartel DP. Predicting effective microRNA target sites in mammalian mRNAs. Elife. 2015; 4

21. Lee CG, McCarthy S, Gruidl M, Timme C, Yeatman TJ. MicroRNA-147 induces a mesenchymal-to-epithelial transition (MET) and reverses EGFR inhibitor resistance. PLoS One. 2014; 9:e84597. [PubMed: 24454732]
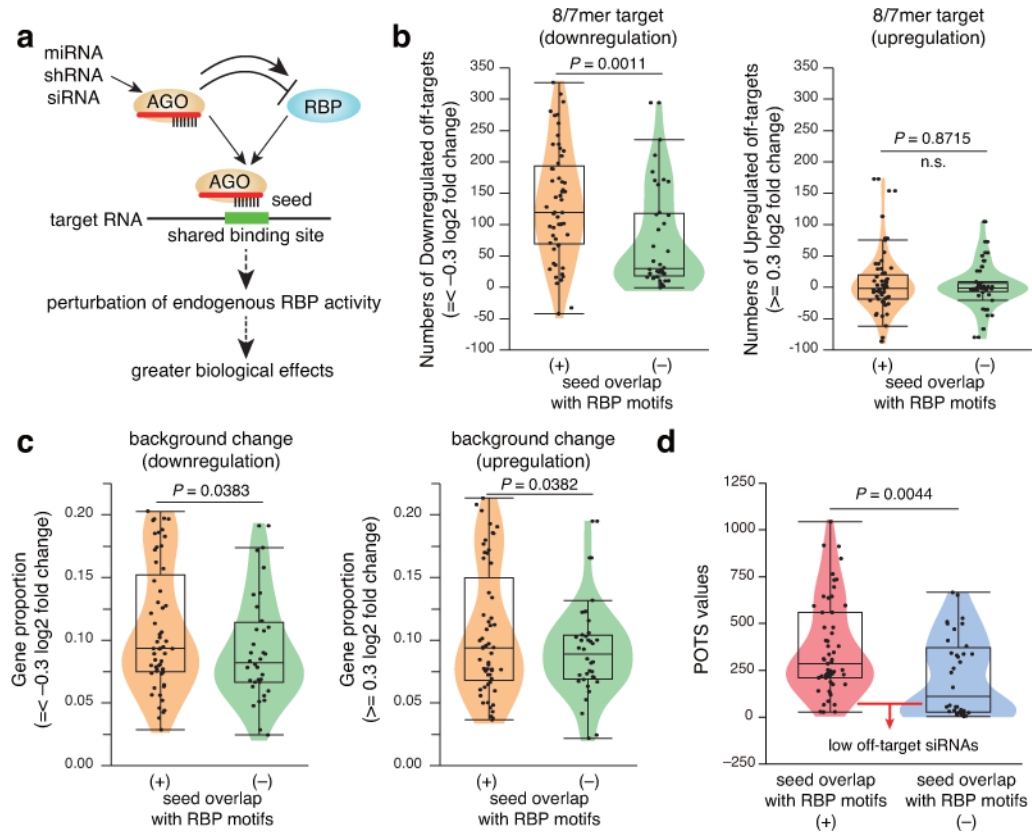
Figure 1. ceRBP concept and impacts of seed-RBP crosstalk on siRNA off-target effects

a, Schematic description of ceRBP concept.

b-d, Effects of overlap between siRNA seed and known RBP motifs on siRNA off-target potency: numbers of downregulated and upregulated off-target genes (b), secondary transcriptome effects, i.e. proportion of downregulated and upregulated background genes (c), and potential off-target score (POTS) (d). Dots represent different transfection experiments using 93 siRNAs. siRNAs with seed overlapping known RBP motifs have higher numbers of downregulated off-target genes (transcripts with 3′ UTRs containing 7- and 8-mer seed-binding sites and -0.3 log2 fold change) over background (b), higher background changes in both down- and upregulation directions (c), and higher POTS values (d). Overlap between siRNA complementary sequences (seed position: 2-8) and RBP motifs more than two bases is considered. Low POTS values (especially < 50) indicate low off-targeting potential. Statistical significance was assessed using two-sided Wilcoxon signed rank test (b,d) or one-sided Welch's t test (c). n.s.: not significant.
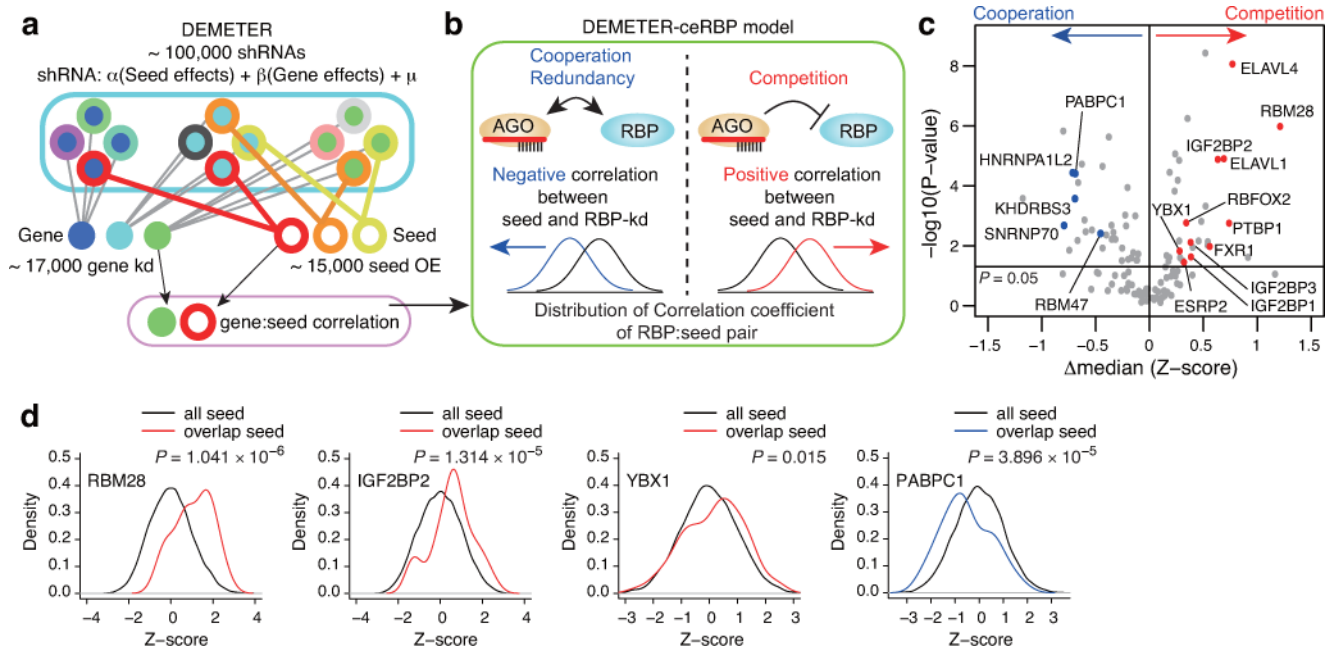
Figure 2. Deconvolution of seed-RBP crosstalk in the Cancer Dependency Map

a, b, Schematic representation of DEMETER model (a) and seed-RBP correlation analysis (b). A image for DEMETER (a) is modified from Tsherniak et al. (2017)[3]. Overlap between 7mer sequences and entire regions of RBP motifs is considered. kd: knockdown.

c, A volcano plot representation of DEMETER-ceRBP analysis showing the median changes of Z-scores of Pearson correlation coefficients (x-axis) and significance (y-axis) for 124 RBPs which overlap at least one seed among 166 RBPs. Distributions of Z-scores for some RBPs are shown in panel (d). P values were calculated by one-sided Kolmogorov–Smirnov (K-S) test. Red and blue dots indicate representative competitive and cooperative relationships, respectively.

d, Distributions of Z-scores of Pearson correlation coefficients for pairs of indicated RBPs and all seeds or motif-overlapping seeds (red: competitive, blue: cooperative). P values were calculated by one-sided Kolmogorov–Smirnov (K-S) test.
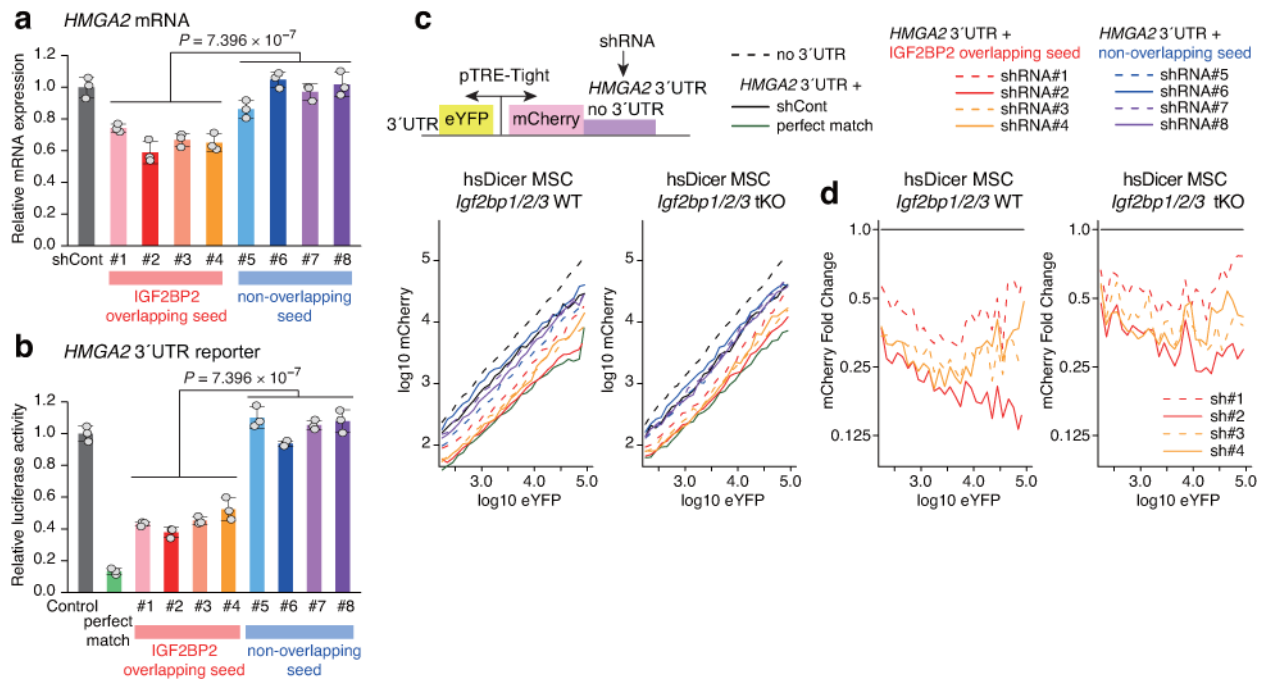
Figure 3. Crosstalk between IGF2BP RBPs and motif-overlapping seeds
a,b, Effects of shRNAs with IGF2BP2 motif-overlapping seeds (#1-#4) and RBP motif non-overlapping seeds (#5-#8) on *HMGA2* mRNA expression (a) and *HMGA2* 3′ UTR luciferase reporter (b) in HEK293T cells (mean and s.d., n = 3, cell cultures). shCont: control shRNA. Two-sided Wilcoxon signed rank test was used to determine significance. **c,d**, Single cell *HMGA2* 3′ UTR reporter assays in *Dicer* KO, human Dicer (hsDicer)-rescued, and *Igf2bp1/2/3* WT or triple KO (tKO) mouse mesenchymal stem cells (MSC) transfected with eYFP and mCherry dual reporter and shRNA plasmids. Panels in (c) show log-log plot of mean mCherry fluorescence across eYFP signal bins, confirming suppressive effects of IGF2BP2 motif-overlapping seeds (#1-#4) relative to RBP motif non-overlapping seeds (#5-#8). Panels in (d) show fold changes of mean mCherry signal. Suppressive effects of motif-overlapping shRNAs are higher in *Igf2bp1/2/3* WT cells relative to tKO cells.
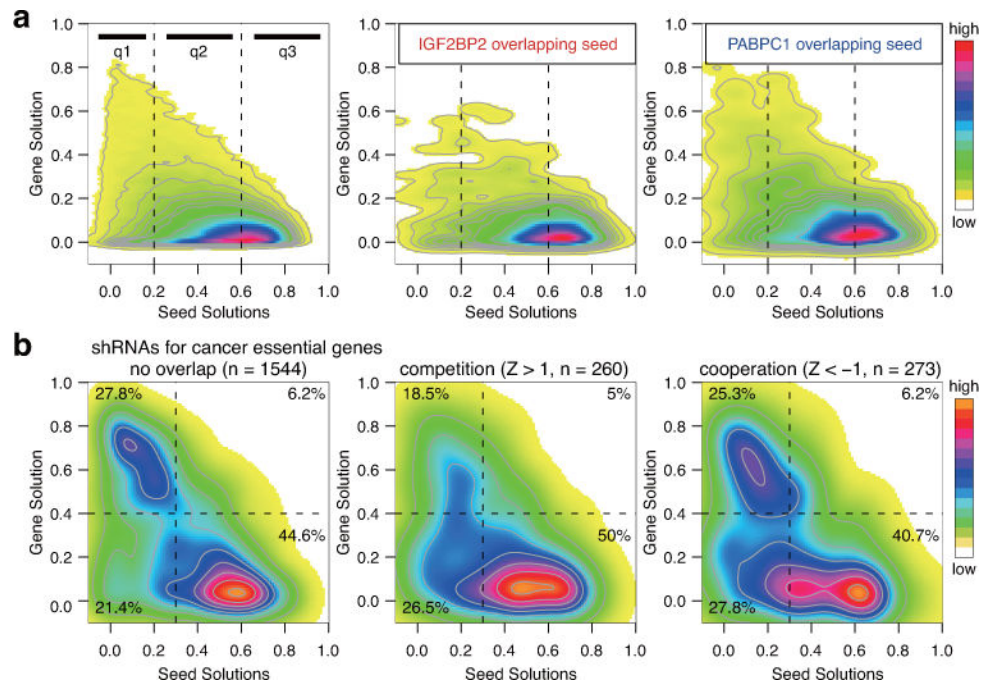
Figure 4. Impacts of seed-RBP crosstalk on shRNA performance in the Cancer Dependency Map
a, Density and contour plots showing distribution of DEMETER gene solution and seed solutions for all shRNAs (~100,000) (left), shRNAs with IGF2BP2 motif-overlapping seeds (middle), and shRNAs with PABPC1 motif-overlapping seeds (right).
b, Distribution of DEMETER gene solution and seed solutions of shRNAs targeting 448 cancer genes differentially required in subsets of 501 cancer cell lines. Plots of shRNAs without RBP overlapping seeds (left) or with competitive (middle) or cooperative (right) seeds are shown. The top-left region corresponds to effective shRNAs.