Supplemental files for "**The mRNA related ceRNA-ceRNA landscape and significance across 20 major cancer types**".

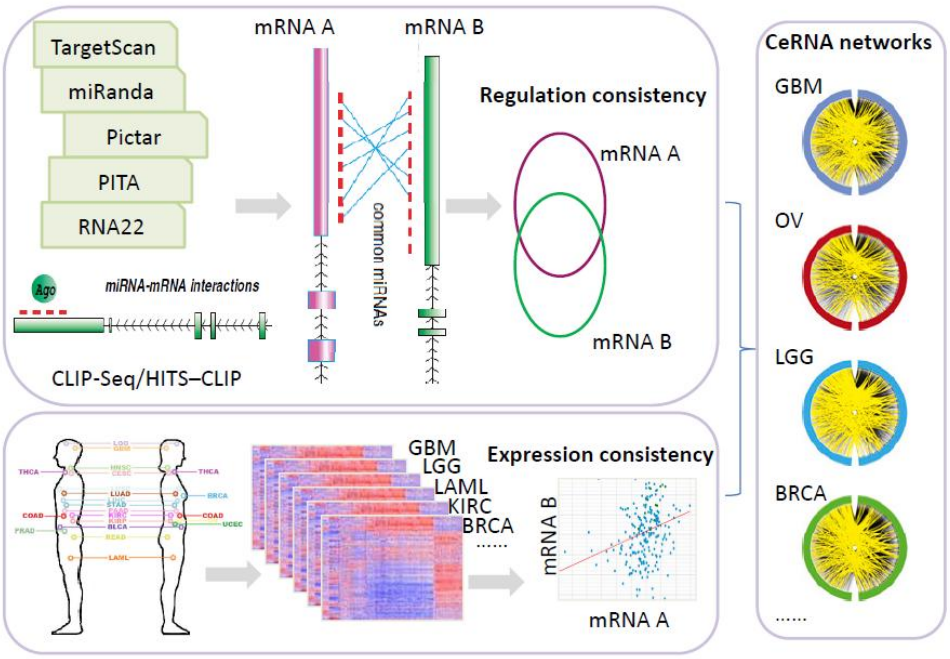

Figure S1. The workflow for constructing and characterizing the miRNA-mediated mRNA related ceRNA network in each cancer.

i) The miRNA-target regulations were identified by integration of the CLIP-seq dataset with five prediction algorithms. And mRNA pairs that are coregulated by miRNAs were identified. ii) The gene expression profiles were collected from the TCGA database the mRNA interactions were identified in each cancer by considering the expression consistency. iii) Cancer-specific ceRNA networks were constructed by assembling all the mRNA related ceRNA-ceRNA interactions.

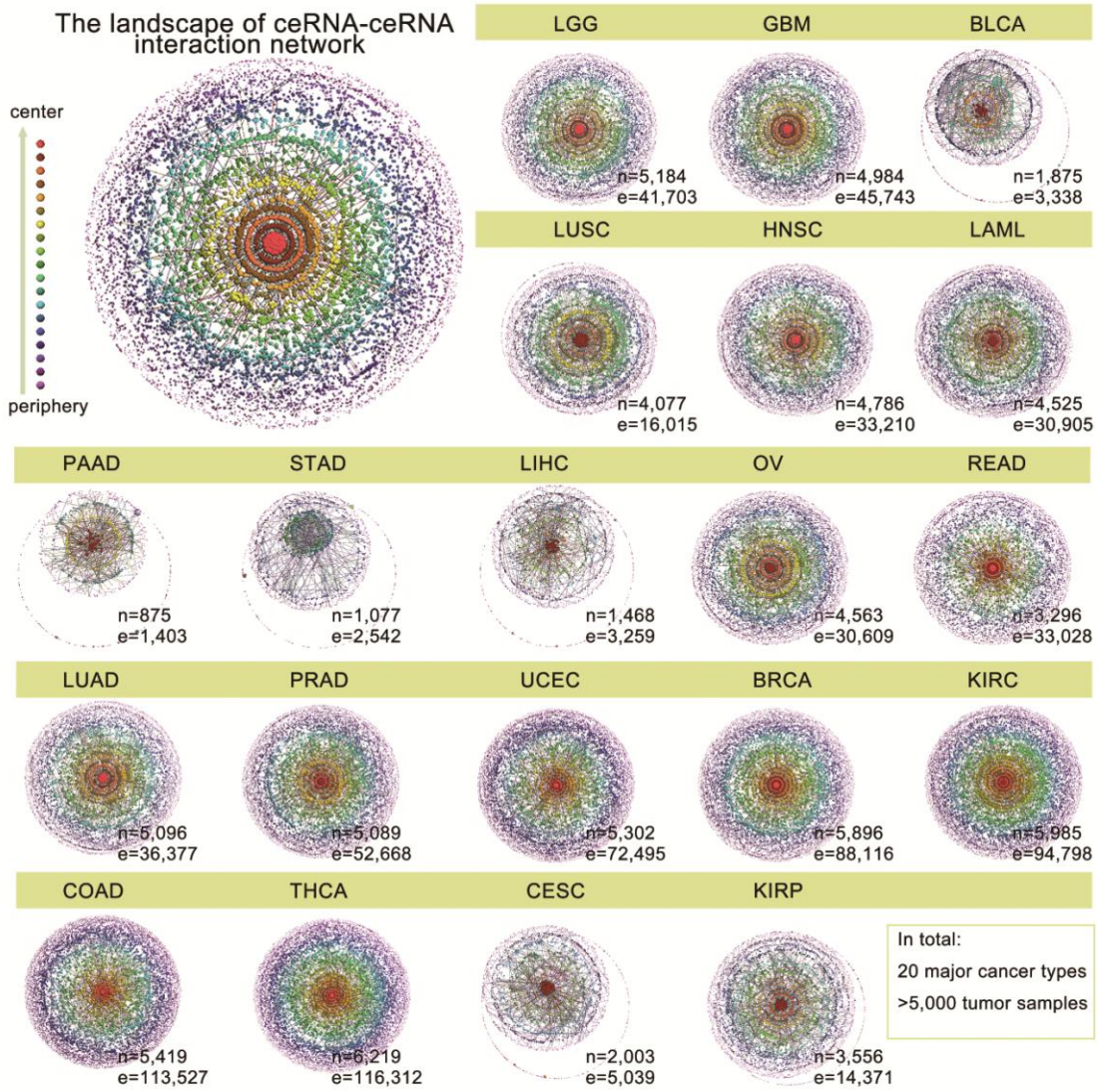

Figure S2. The landscape of ceRNA-ceRNA interaction networks across 20 types of cancer.

Its graphic visualization uses nodes to represent individual ceRNAs and edges to represent miRNA-mediated RNA-RNA interactions. Nodes near the center of the graph are contained within more tightly regulated, dense subnetworks. The color bands which include nodes with similar connectivity, have a size increases with the distance from the center. The networks are visualized with the Lanet plugin in the network workbench.

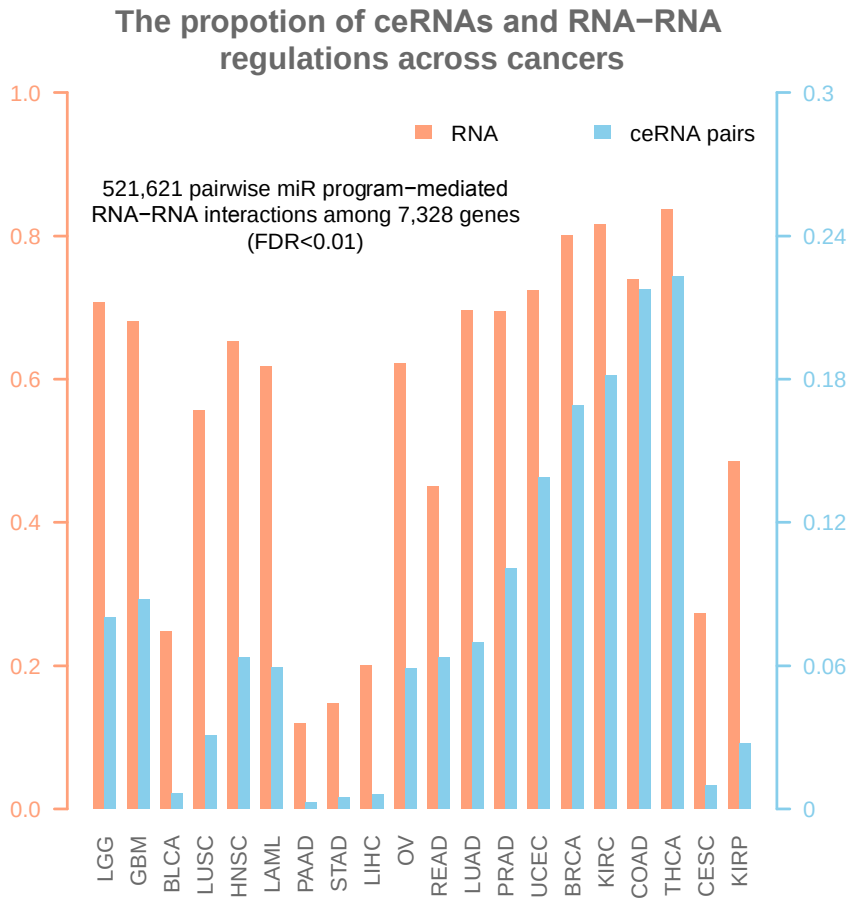

Figure S3. The percentage of nodes and edges of the ceRNA-ceRNA networks across 20 types of cancer.

The left y-axis represents the percentage of ceRNAs in each ceRNA network, corresponding to the orange bars. The right y-axis represents the percentage of interactions in each network, corresponding to the blue bars.

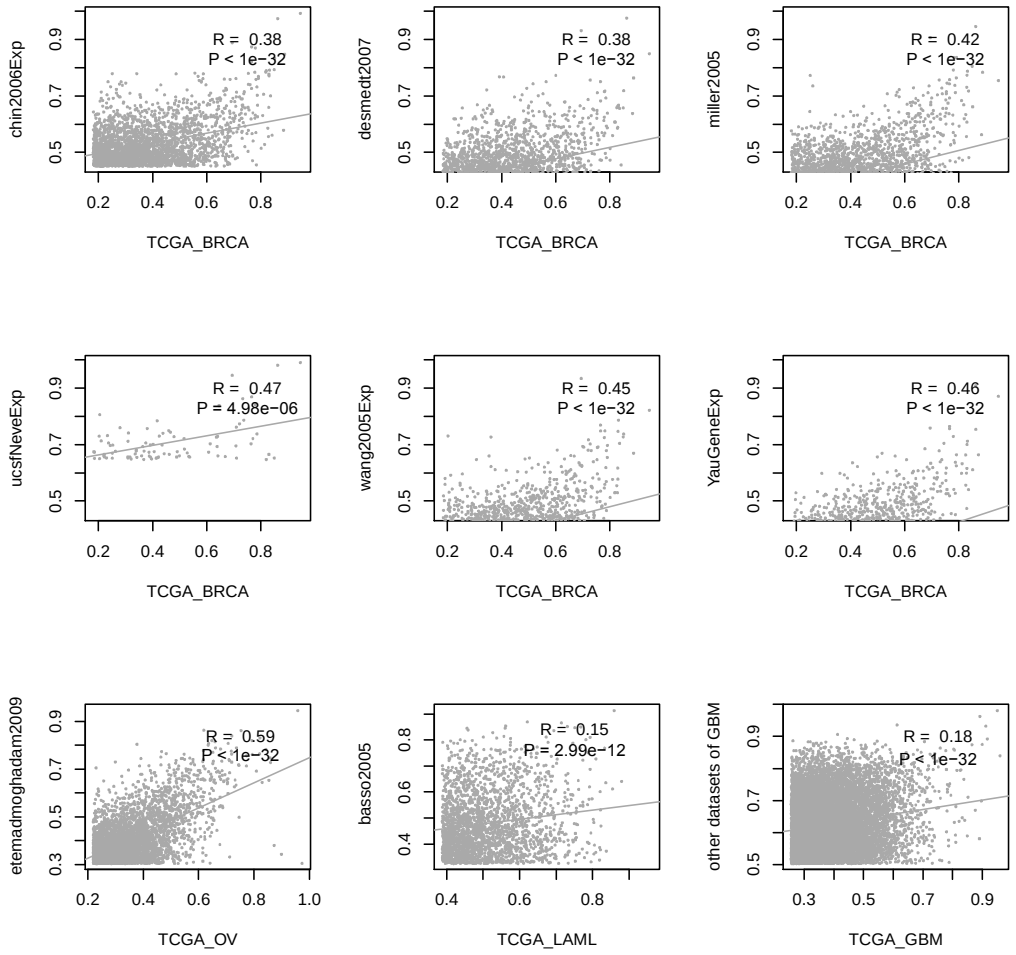

Figure S4. The ceRNA networks were robust in independent datasets.

The expression consistency of ceRNA-ceRNA interactions were significantly correlated with each other in independent datasets. These independent datasets were obtained from published literature with pubmed IDs as follow: 17157792, 17545524, 16141321, 17157791, 15721472, 20946665, 19193619, 15778709, 24194606.

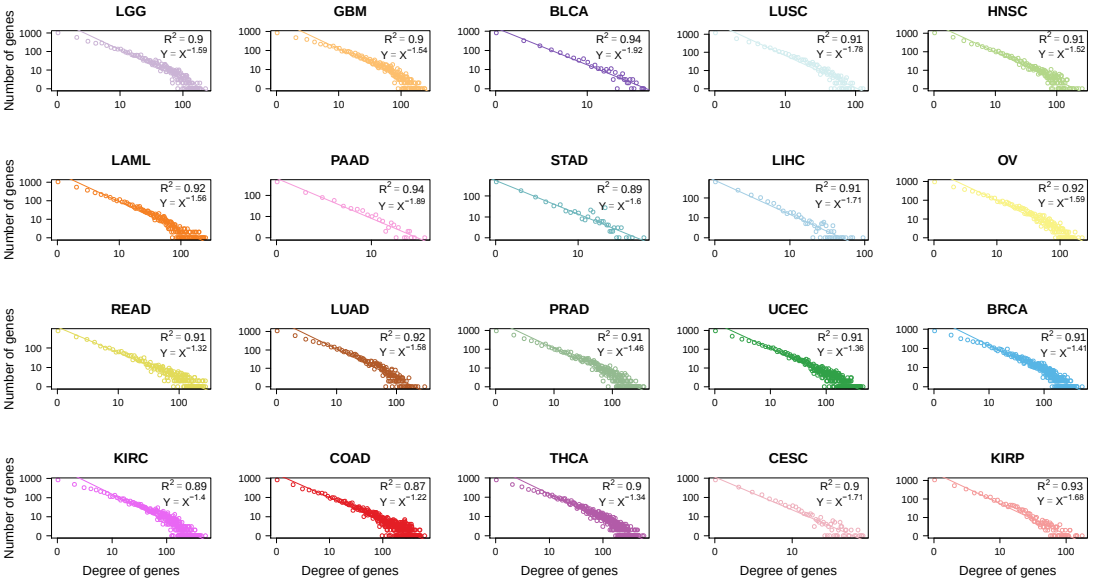

Figure S5. The Degree distribution of the ceRNA networks across 20 types of cancer.

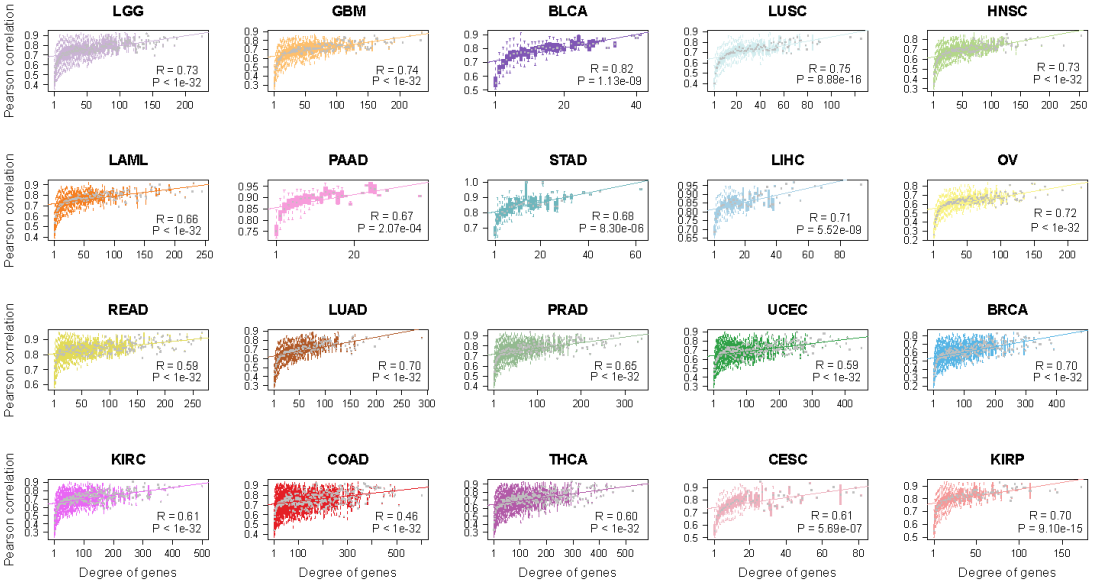

Figure S6. Hub ceRNAs are more coexpressed with their neighbors than others.

The correlation between expression of ceRNAs and the total expression of their ceRNA regulators is plotted as a function of the number of its ceRNA regulators; genes at the center of the ceRNA network are regulated by hundreds of ceRNA regulators and are significantly correlated with their total expression.

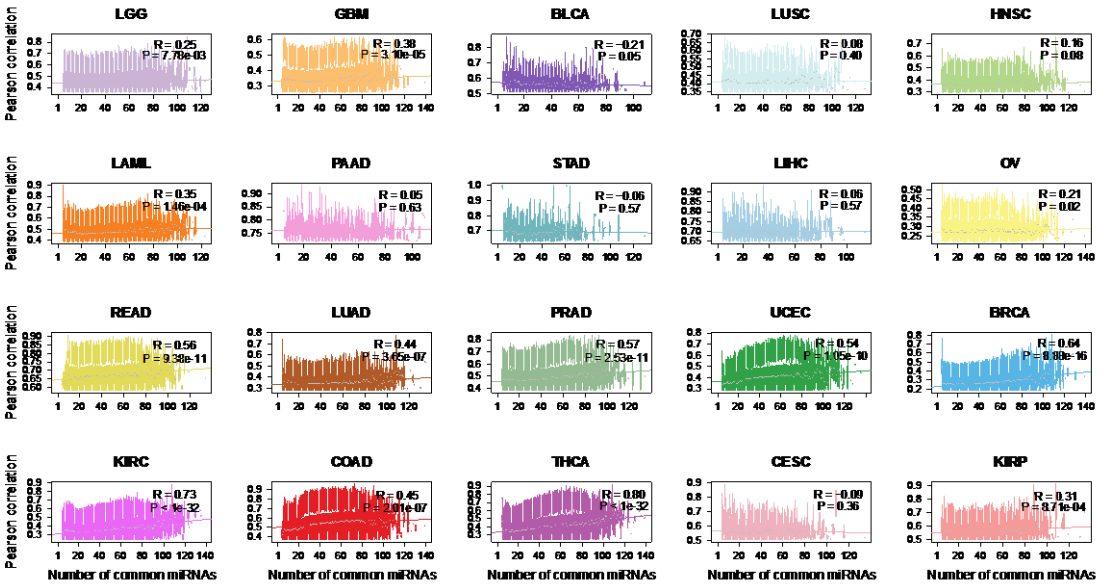

Figure S7. Coexpression of ceRNAs in the network increases with the number of common miRNAs.

The ceRNA interactions were grouped by the number of miRNAs they share, and then the correlation coefficient of expression was shown as boxplot in each group.

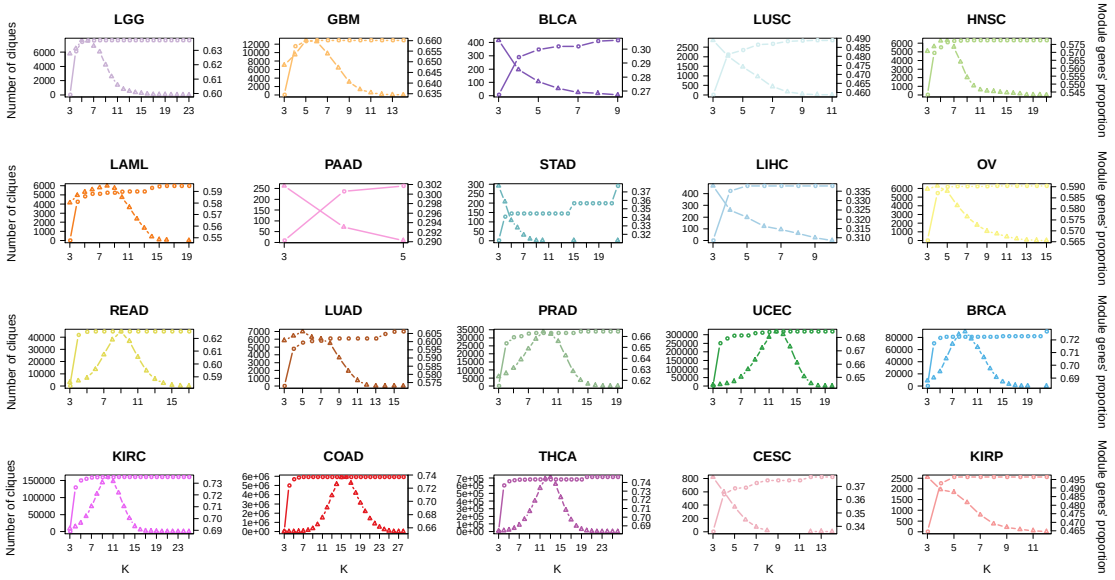

Figure S8. Number of cliques at different k-values and cumulative ratios of ceRNAs in cliques with k-values are not bigger than k.

The left y-axis represents number of cliques under different k-values, corresponding to the triangle line. The right y-axis represents cumulative ratios of ceRNAs in cliques, corresponding to the circle line.

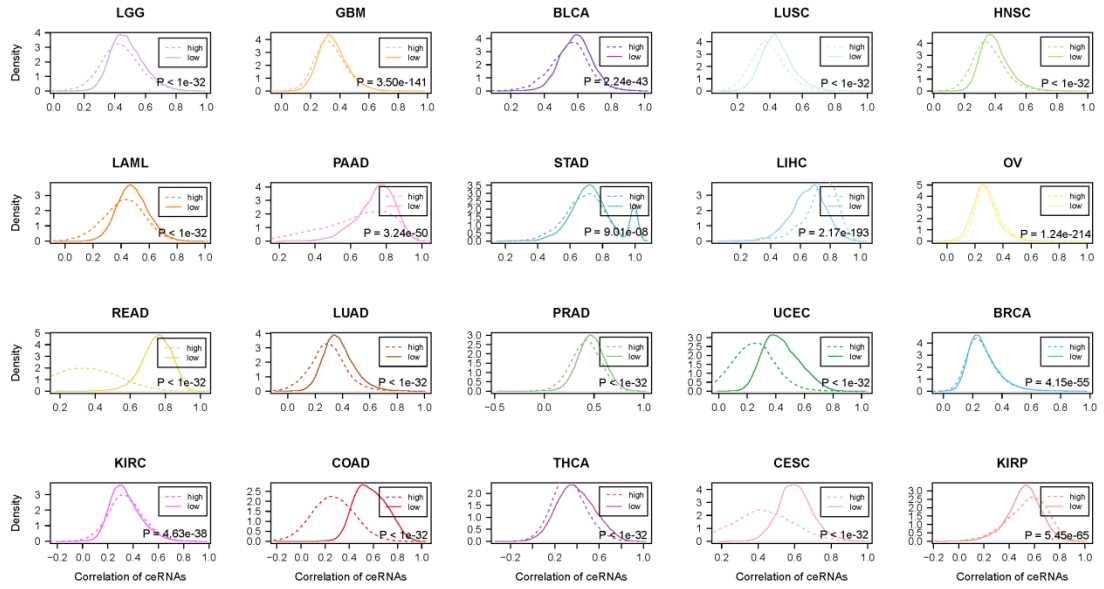

Figure S9. ceRNAs were strongly coexpressed in Dicer-low expressed groups.

The solid lines represent the distribution of correlation coefficient of ceRNA pairs in the samples that are low expressed of Dicer; and the dash lines represent the distribution of correlation coefficient of ceRNA pairs in the samples that are high expressed of Dicer. The differences of the two distributions were tested by ranksum test.

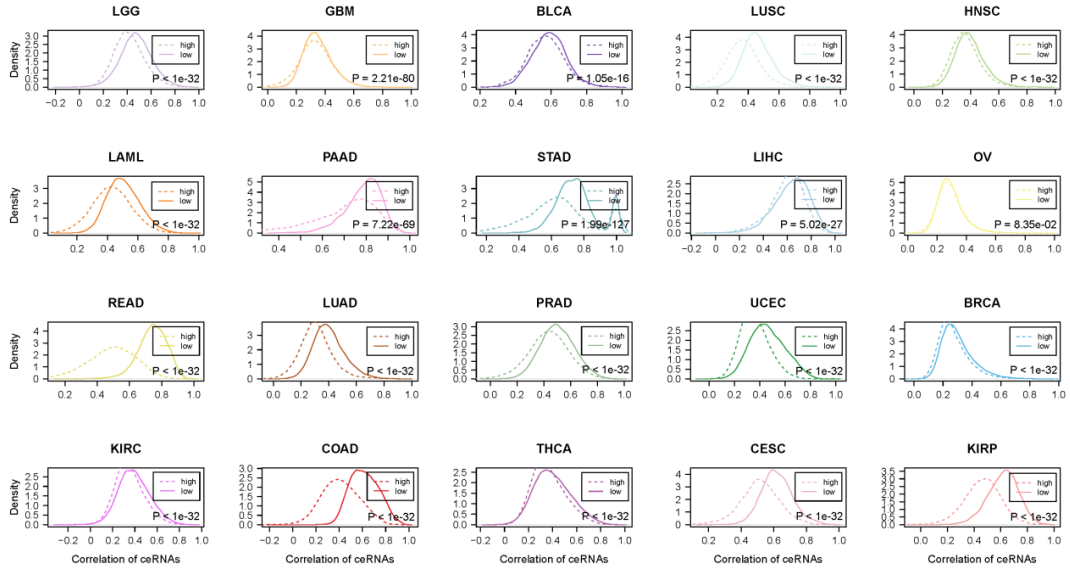

Figure S10. ceRNAs were strongly coexpressed in Drosha-low expressed groups.

The solid lines represent the distribution of correlation coefficient of ceRNA pairs in the samples that are low expressed of Drosha; and the dash lines represent the distribution of correlation coefficient of ceRNA pairs in the samples that are high expressed of Drosha. The differences of the two distributions were tested by ranksum test.

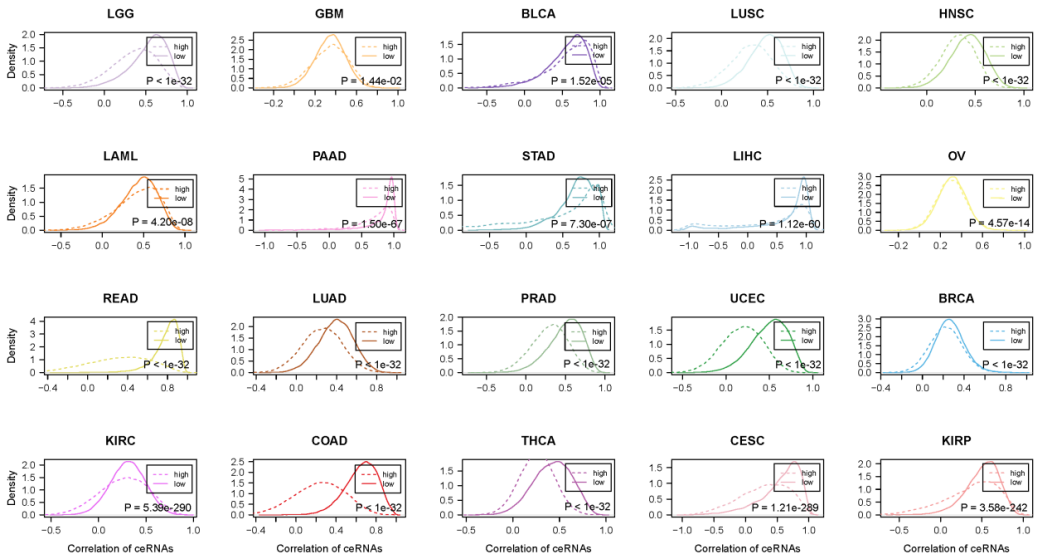

Figure S11. ceRNAs were strongly coexpressed in Dicer/Drosha-low expressed groups.

The solid lines represent the distribution of correlation coefficient of ceRNA pairs in the top 30% samples that are low expressed of Dicer and Drosha; and the dash lines represent the distribution of correlation coefficient of ceRNA pairs in the top 30% samples that are high expressed of Dicer and Drosha. The differences of the two distributions were tested by ranksum test.

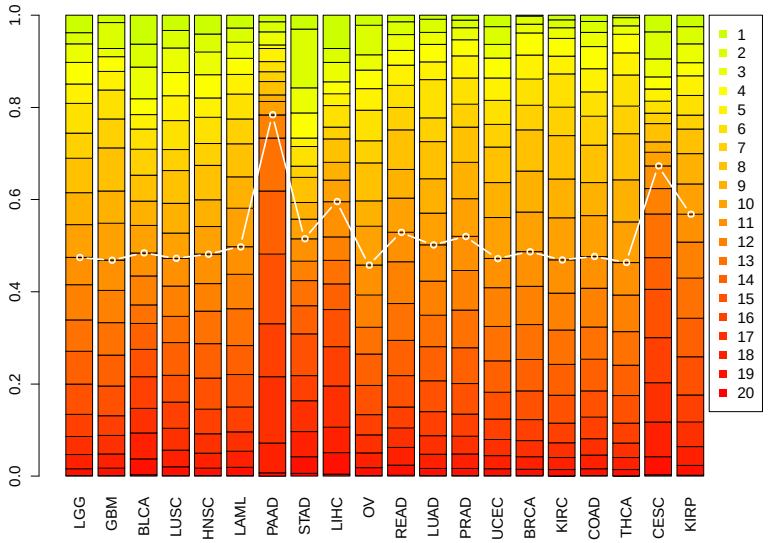

Figure S12. ceRNA hubs (Top 15%) retained their high degree across cancers.

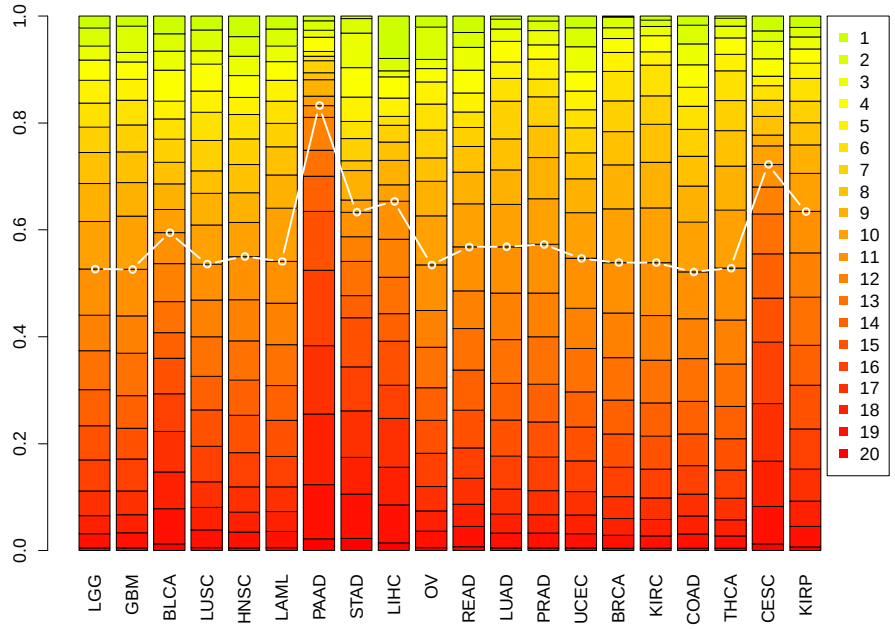

Figure S13. ceRNA hubs (Top 20%) retained their high degree across cancers.

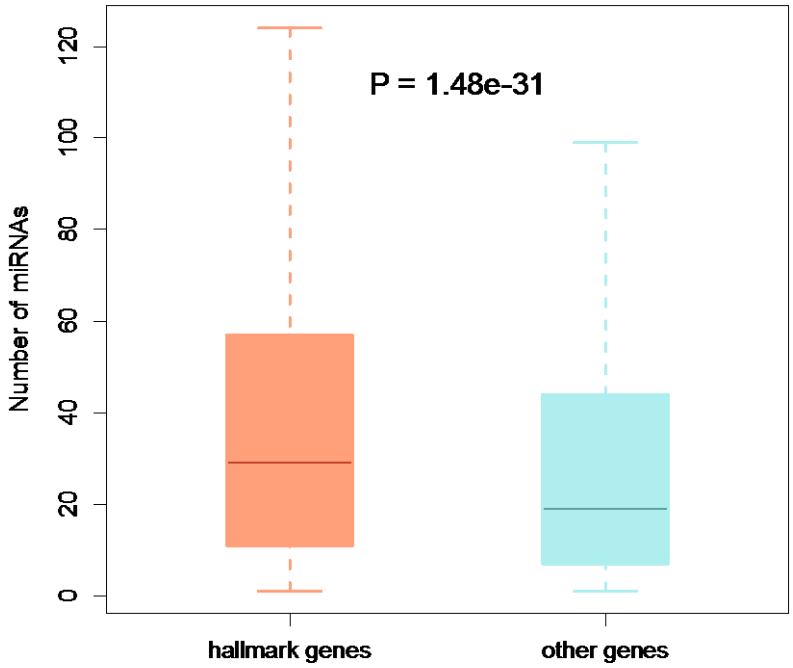

Figure S14. The hallmark genes were regulated by more miRNAs than other genes.

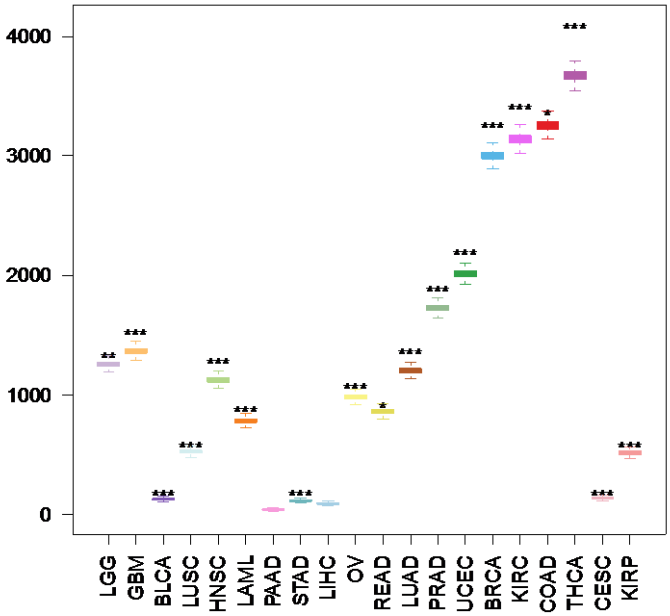

Figure S15. The hallmark associated ceRNA networks are much denser than expected by chance.

The number of edges in the random hallmark associated ceRNA networks were plot as a box in each cancer, and the real number was marked as a start.

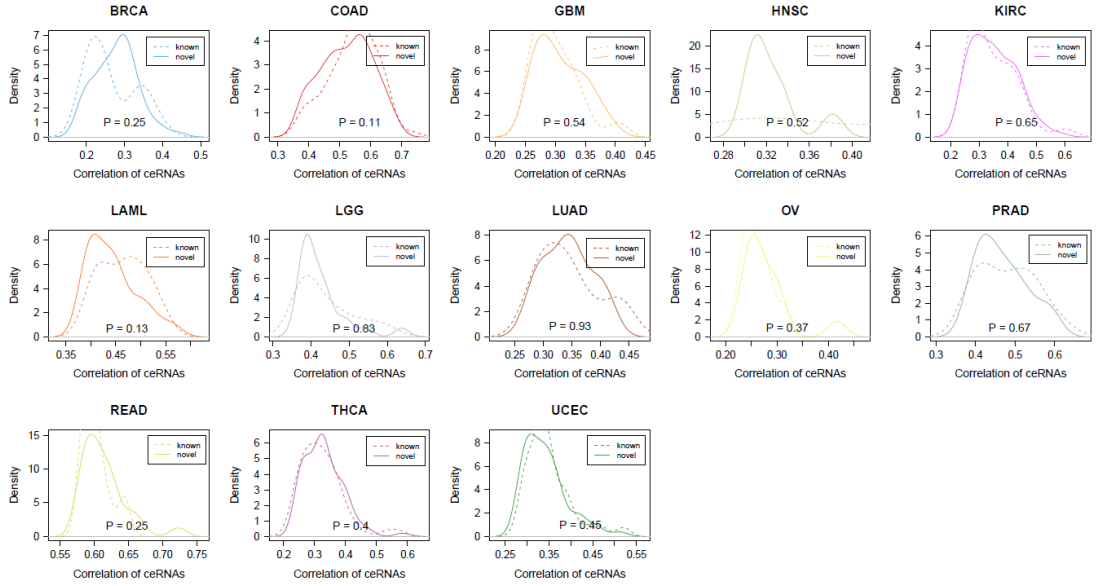

Figure S16. The expression similarity of the predicted and known ceRNAs of PTEN are higher than random conditions. The distribution were tested by KS-test.

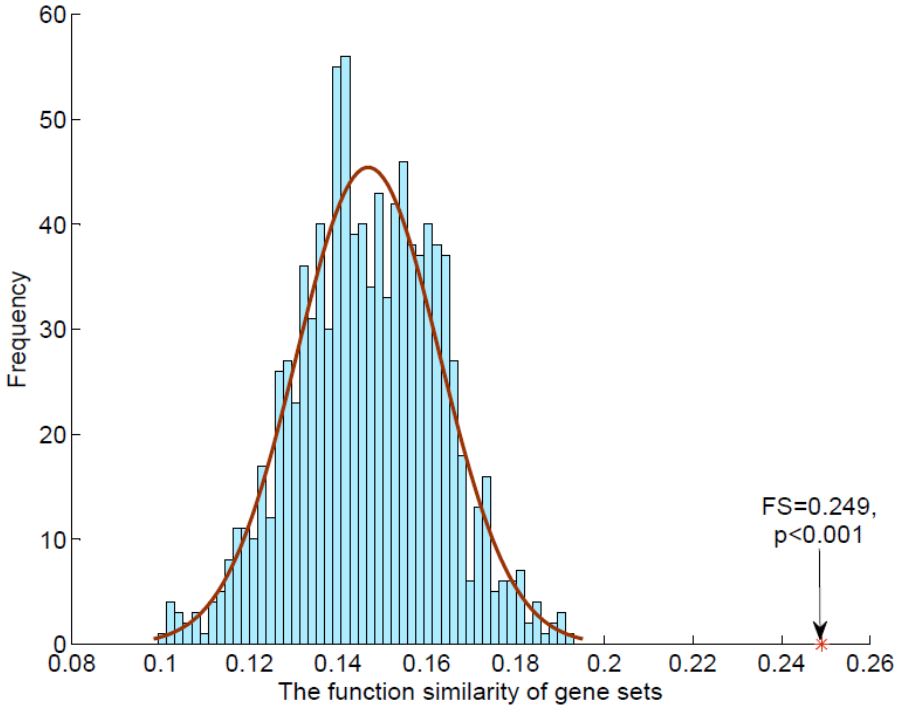

Figure S17. The function similarity of the predicted and known ceRNAs of PTEN are higher than random conditions.

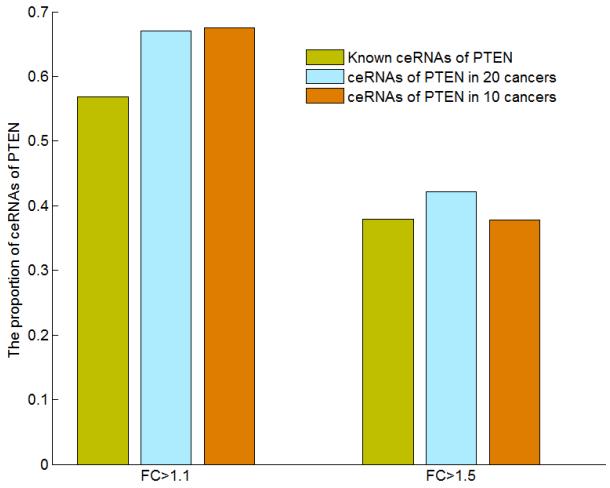

Figure S18. The expression of ceRNAs of PTEN were changed when PTEN overexpress in U87 cell lines.

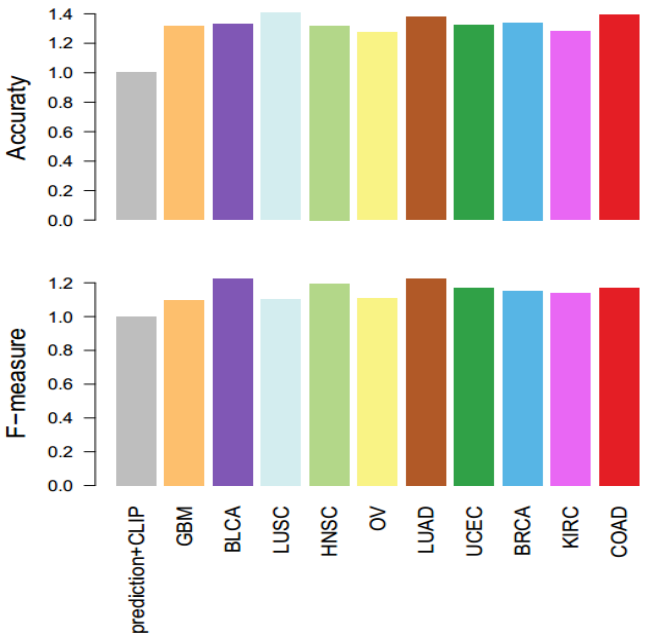

Figure S19. The accuracy and F-score of the integration method were larger than that of Ago CLIP-supported miRNA interactions.

The accuracy was the proportion of miRNA-gene regulations that are inlcuded in these experimentally verified databases in each cancer. And the F-score was calculated as 2*precision*recall/(precision+recall). The accuracy and F-score of CLIP-supported miRNA-regulations were normalized to one.

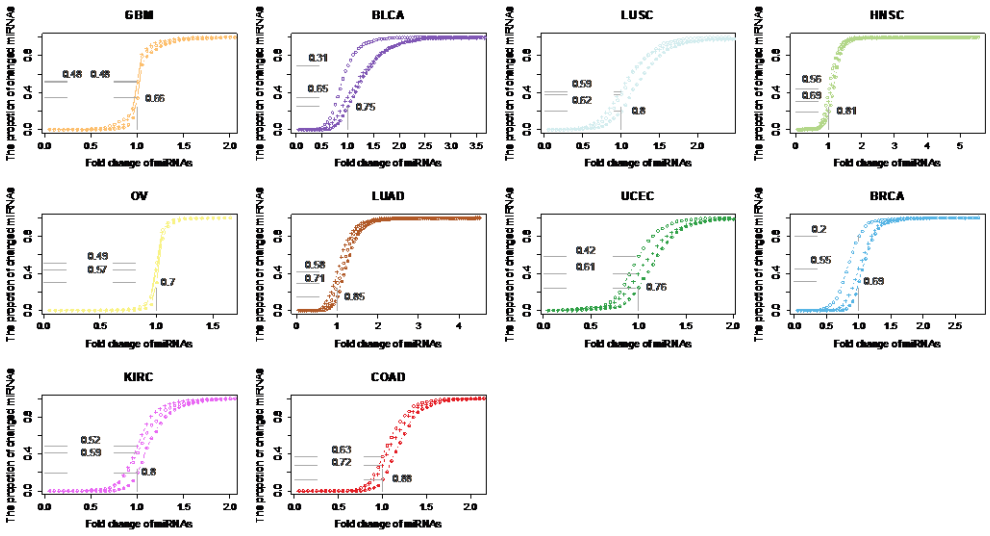

Figure S20. The expression of miRNAs were regulated by Dicer and Drosha in various types of cancers.

The cumulative distribution of fold change of miRNA expression in Dicer or Drosha low and high expressed samples. The line with ‘o’ shows the distribution of fold changes in Dicer-low and Dicer-high samples. The line with ‘+’ shows the distribution of fold changes in Drosha-low and Drosha-high samples. The line with ‘+o’ shows the distribution of fold changes in Dicer/Drosha-low and Dicer/Drosha-high samples.

Table S1. The genome-wide gene expression profiles used in our current study.

Cancer	Sample	Gene	Platform	Node	Edges
LGG	205	20,051	RNA-seq(Hiseq)	5,184	41,703
GBM	403	17,813	MicroArray	4,984	45,743
BLCA	96	20,500	RNA-seq(Hiseq)	1,815	3,338
LUSC	220	20,500	RNA-seq(Hiseq)	4,077	16,015
HNSC	303	20,500	RNA-seq(Hiseq)	4,786	33,210
LAML	173	20,500	RNA-seq(Hiseq)	4,525	30,905
PAAD	41	19,726	RNA-seq(Hiseq)	875	1,403
STAD	58	20,357	RNA-seq(GA)	1,077	2,542
LIHC	54	19,777	RNA-seq(Hiseq)	1,468	3,259
OV	558	17,813	MicroArray	4,563	30,609
READ	71	20,500	RNA-seq(Hiseq)	3,296	33,028
LUAD	355	20,500	RNA-seq(Hiseq)	5,096	36,377
PRAD	179	20,067	RNA-seq(Hiseq)	5,089	52,668
UCEC	333	20,500	RNA-seq(Hiseq)	5,302	72,495
BRCA	822	20,500	RNA-seq(Hiseq)	5,896	88,116
KIRC	470	20,500	RNA-seq(Hiseq)	5,985	94,798
COAD	192	20,500	RNA-seq(Hiseq)	5,419	113,527
THCA	470	20,117	RNA-seq(Hiseq)	6,219	116,312
CESC	99	19,982	RNA-seq(Hiseq)	2,003	5,039
KIRP	101	20,043	RNA-seq(Hiseq)	3,556	14,371

Table S2. The miRNA target sites were likely to localize on 3'UTR.

	CLIP-Seq	Genes with expression
3'UTR	607,232 (68.21%)	578,804 (69.05%)
CDS	242,701 (27.26%)	222,920 (26.59%)
5'UTR	40,293 (4.53%)	36,527 (4.36%)

Table S3. The conserved ceRNA modules across cancers. (Table S3.xlsx)

Table S4. The cancer specific ceRNA modules. (Table S4.xlsx)

Table S5. The ceRNAs of PTEN across 20 types of cancer. (Table S5.xlsx)

Table S6. The proportion of ceRNA pairs that are co-localization or co-regulation across 20 types of cancer.

Cancer	Co-localization	Co-regulation	Union
LGG	0.39%	0.09%	0.47%
GBM	0.06%	0.07%	0.13%
BLCA	3.83%	3.42%	7.25%
LUSC	0.71%	1.27%	1.74%
HNSC	0.27%	0.61%	0.86%
LAML	0.98%	0.97%	1.79%
PAAD	9.05%	0.57%	9.48%
STAD	13.45%	13.02%	19.59%
LIHC	5.52%	0.92%	6.29%
OV	0.12%	0.15%	0.27%
READ	0.81%	0.08%	0.90%
LUAD	0.41%	0.79%	0.99%
PRAD	0.49%	0.32%	0.81%
UCEC	0.19%	0.35%	0.51%
BRCA	0.18%	0.39%	0.45%
KIRC	0.15%	0.25%	0.37%
COAD	0.70%	0.10%	0.80%
THCA	0.15%	0.21%	0.28%
CESC	3.87%	3.97%	6.53%
KIRP	0.89%	0.32%	1.18%

Dataset S1. The hallmark associated ceRNA networks across 20 types of cancer.

Supplemental methods

Construction of the ceRNA networks in individual cancer types

Firstly, a hypergeometric test is executed for each possible gene pairs separately. For each given gene pair of gene A and B, we identified the common miRNA regulated them (A∩B). The subset is required to have at least O_{min} miRNAs. And then the probality P for gene A and B is calculated according to

$$P = 1 - F(x \mid N, K, M) = 1 - \sum_{t=0}^{x-1} \frac{\binom{K}{t}\binom{N-K}{M-t}}{\binom{N}{M}}$$

where N is the number of all miRNAs, K and M is the total number of miRNAs regulated gene A and B, x is the common miRNA number between these two genes. Only gene pairs that regulated by three common miRNAs were analyzed in our study. All P-values were subject to false discovery rate (FDR) correction. Candidate gene pairs with FDR less than 0.01 were used for subsequent analyses.

Next, integrated with gene expression profiles in all individual cancer types, we identified the ceRNA pairs in specific tumor type. To explore the active ceRNA pairs in individual cancer, we computed the correlation coefficient (R) of each candidate ceRNA pairs identified above. All the candidate gene pairs with R>0 and p-adjusted<0.05 were identified as ceRNA-ceRNA interactions. After assembling all significant ceRNA pairs, we generated the ceRNA network for each cancer type. A node represents a gene, and two nodes are connected if they were coregulated by miRNAs and co-expressed in this cancer.

Identification of co-localized and co-regulated gene pairs

The protein-coding genes within 5 kb of each other were regarded as being co-localized gene pairs. In addition, we downloaded the TF-gene regulations from the ChIPBase database (1) and then used linear regression to identify the active TF-gene regulations in each cancer (FDR<0.01). As in an earlier study (2), two overlap ratios were calculated for a protein coding gene A and another protein coding gene B with

different numbers of TFs: the proportion of TFs regulating A that were also regulating B (r_{AB}), and the proportion of genes regulating B that were also regulating A (r_{BA}). We chose the formula $r=(r_{AB}*r_{BA})^{0.5}$ to describe the degree of coregulation. The gene pairs with an r greater than 0.8 were regarded as co-regulated.

The function similarity between the novel ceRNAs and known ceRNAs of PTEN

In order to estimate the function correlation between the novel ceRNAs and the knonw ceRNAs of PTEN, we calcualted the function similarity for these two ceRNA sets based on the GS^2, which quantifies the similarity of the Gene Ontology (GO) annotations among two gene sets. Moreover, the significance of functional similarity was caluculated by randomization test. We randomly chose the same number of genes as the novel ceRNA sets and recomputed the function similarity. This process was repeated 1,000 times. And the P-value is the fraction of function similarity in random conditions, which is larger than that of real one.

Sensitivity correlation for ceRNA-ceRNA interactions

For each ceRNA interaction between gene A and B, we then computed the maximum difference between the Pearson and partial correlation coefficients according to each shared miRNA and defined it sensitivity correlation (S) :

$$S_{AB} = \max(\mathrm{corr}(\mathrm{mRNA_A}, \mathrm{mRNA_B}) - \mathrm{corr}(\mathrm{mRNA_A}, \mathrm{mRNA_B}|\mathrm{miRNA}))$$

In addition, we permutated the miRNA expression to evaluate the significance of the S score. This process was repeated 1,000 times. The P-value is the fraction of S for all random conditions, which is greater than the real one, which were further adjusted by the BH-method.

Supplemental Text S1

Benchmark analyses

The regulation and expression consistency is commonly used to investigate the ceRNA-ceRNA regulations. To evaluate whether this two-steps approach is valuable to learn the ceRNA-ceRNA network from an mRNA-mRNA correlation network, we first defined a competing activity score (CAS) for each ceRNA-ceRNA interactions as follow:

$$\mathrm{CAS} = -\log(\mathrm{p}_{\mathrm{miRNA}}) - \log(\mathrm{p}_{\mathrm{exp}})$$

where $\mathrm{p}_{\mathrm{miRNA}}$ evaluates the statistical significance of two mRNAs share the common miRNAs and $\mathrm{p}_{\mathrm{exp}}$ evaluates the significance of expression correlation of two mRNAs. The higher the activity score is, the more strong competion between these two mRNAs is.

The performance of the CAS measurement was estimated by comparing gold standard ceRNA-ceRNA interactions with random interactions. PTEN-associated ceRNAs (n=607) were collected from previous studies (3-11) and considered as gold standard ceRNA-ceRNA interactions. For each PTEN-ceRNA interaction, 100 random CASs were generated by randomly shuffling the expression profiles and miRNA-regulations. As a result, we found that the CAS values of known PTEN-related ceRNA pairs were significantly higher than random scores across 20 types of cancer (**Figure S21**). Moreover, the CAS scores of PTEN-related ceRNA pairs were also significantly higher than those of coexpressed gene pairs (**Figure S22**). These results suggest that the CAS index can give high levels of precision to distinguish positive ceRNA interactions from negative ones.

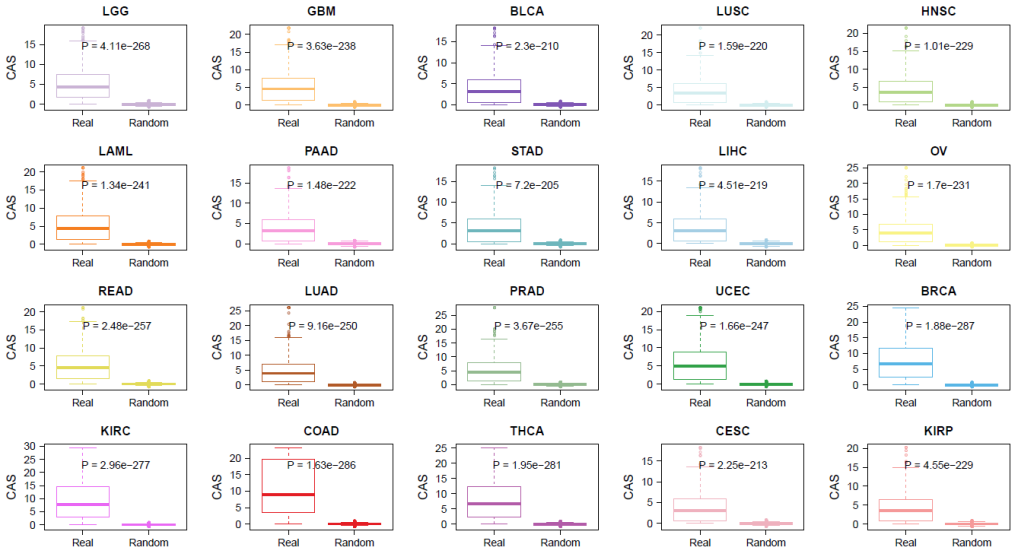

Figure S21. Box plot comparison of competing activity scores between PTEN-associated ceRNAs and randoms. The scores of PTEN-associated ceRNAs were significantly higher than random scores across 20 types of cancers. Significant P values were calculated by the Wilcox-ranksum test.

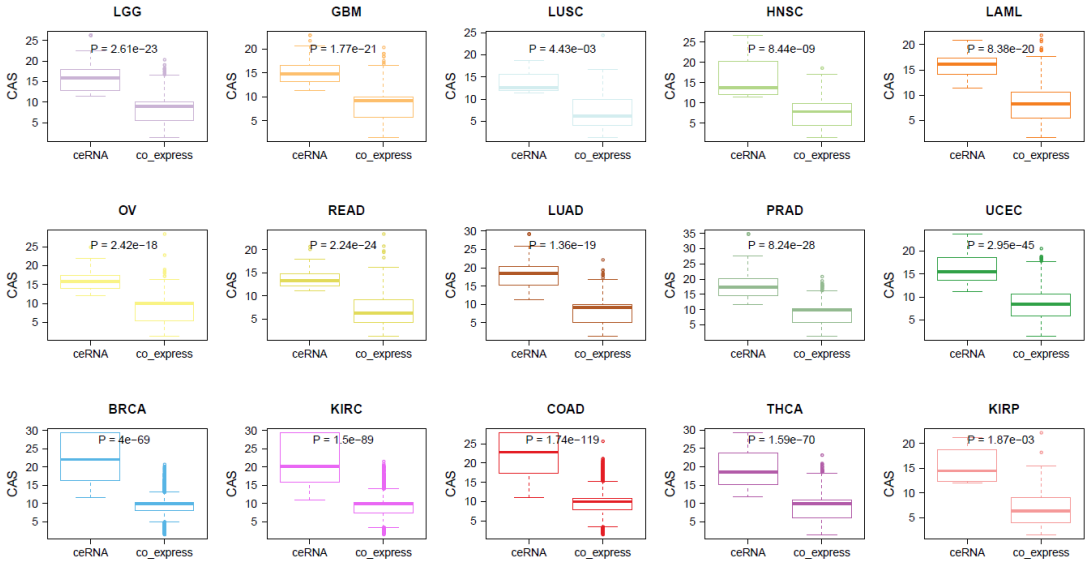

Figure S22. Box plot comparison of competing activity scores between PTEN-associated ceRNAs and coexpression pairs of PTEN. The scores of PTEN-associated ceRNAs were significantly higher than coexpression pairs across 20 types of cancers. Significant P values were calculated by the Wilcox-ranksum test.

Our method to identify ceRNA interactions was further evaluated based on operating characteristic (ROC) curves. The negative training examples of ceRNA pairs were detected based on the negative miRNA-target interactions obtained from Bandyopadhyay et al. (12). Two candidate mRNAs were identified as negative

ceRNA interactions if they interacted with a common miRNA. The gold standard positive and negative ceRNA interactions were further merged into a list ranked by CAS in descending order for which dynamic thresholds (ranging from minimum to maximum of competing activity scores) were used as cut-off points. Receiver operating characteristic (ROC) curves were ploted by using the CASs. As a result, high AUC values (from 0.743 to 0.914) for ROC curves across 20 cancers (**Figure 23**). These results validate that this two-steps approach is valuable to learn the ceRNA-ceRNA network and the ceRNA networks across cancers can be used to understand the biological mechanism of cancers.

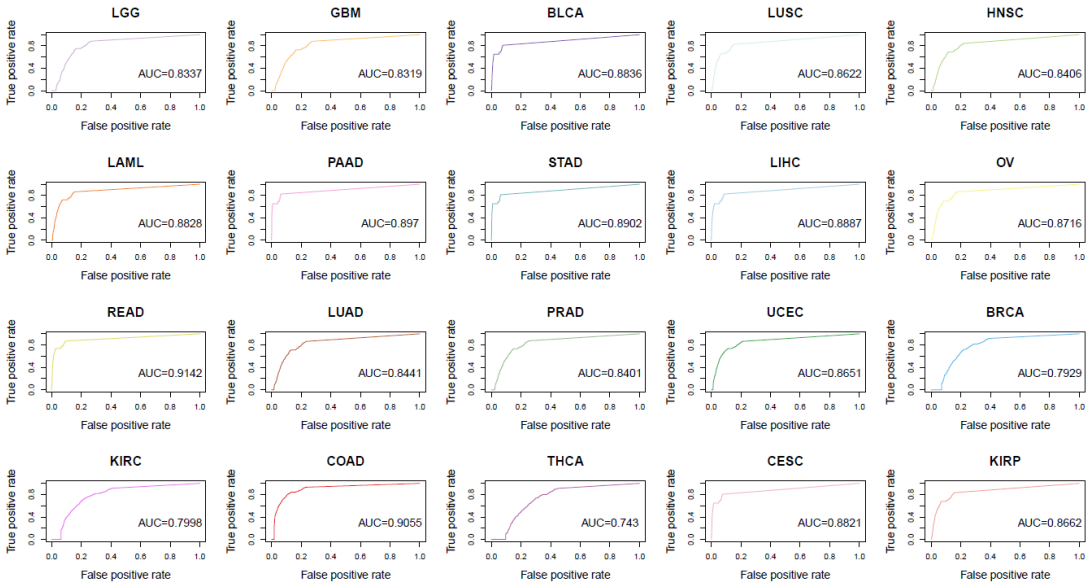

Figure S23. The ROC curves used to distinguish PTEN-associated ceRNAs from negative ceRNAs, based on the competing activity score.

Supplemental Text S2

Considering other scenarios that may lead to coexpression between mRNAs, such as co-localization and co-regulation by same transcription factors, we found that there are only 0.13%-19.59% ceRNA pairs were co-localization or co-regulated by TFs (**Table S6**). Analyzing the ceRNA networks after filtering those co-localized or co-regulated pairs, we obtained the similar topological and functional landscapes of ceRNAs networks across human cancers. Analysis of the ceRNA networks, we obtained the similar topological and functional landscapes of ceRNAs networks

across human cancers (**Figure S24-S31**). The ceRNA networks also show scale-free and modular structures and the ceRNA pairs were strongly coexpressed in Dicer/Drosha-low expressed groups. In addition, we found that in the Dicer/Drosha-low expressed groups. These results further evidence that the structures of the ceRNA networks and most of the results obtained in our study are robust.

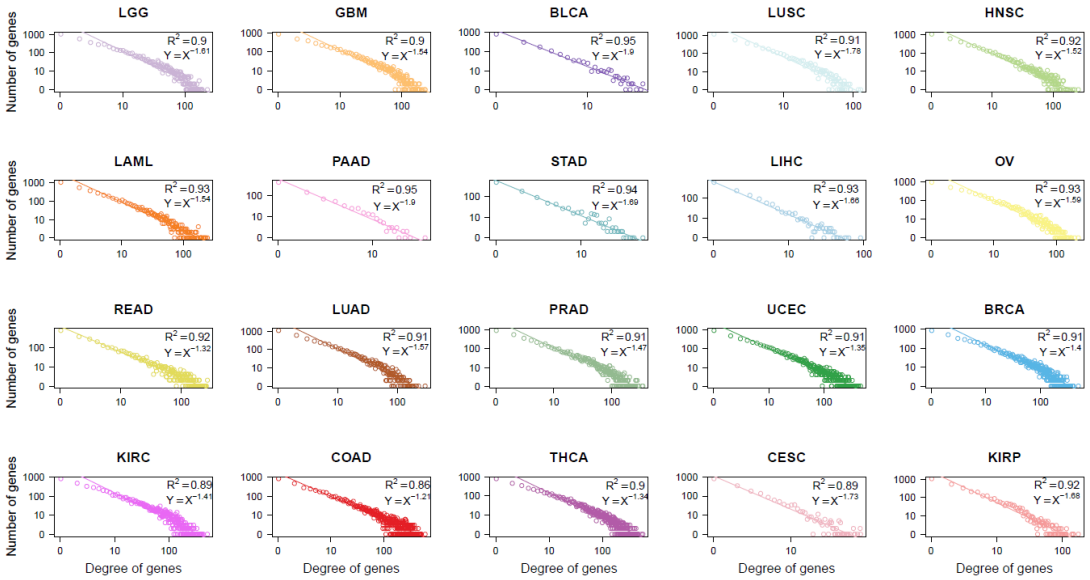

Figure S24. The degree distributions of the ceRNA networks across cancers.

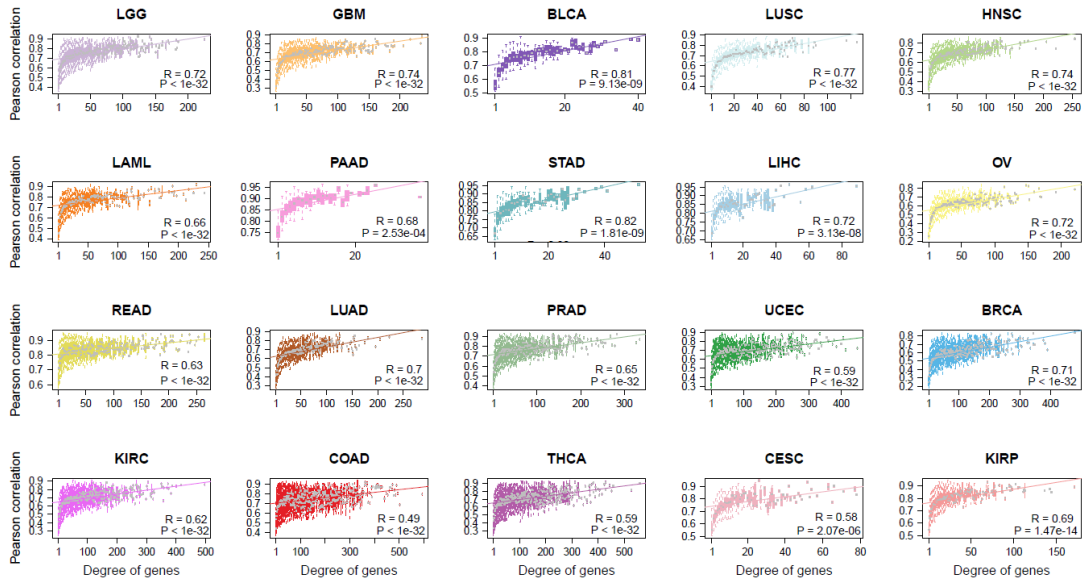

Figure S25. Hub ceRNAs are more coexpressed with their neighbors than others. The correlation between expression of ceRNAs and the total expression of their ceRNA regulators is plotted as a function of the number of its ceRNA regulators; genes at the center of the ceRNA network are regulated by hundreds of ceRNA regulators and are significantly correlated with their total expression.

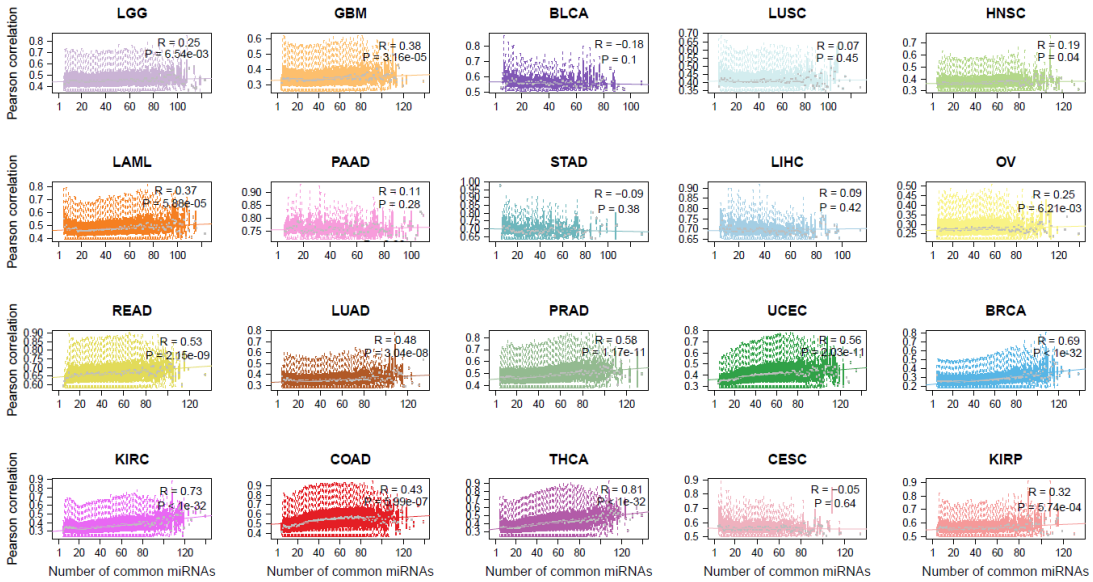

Figure S26. Coexpression of ceRNAs in the network increases with the number of common miRNAs. The ceRNA interactions were grouped by the number of miRNAs they share, and then the correlation coefficient of expressions were shown as boxplot in each group.

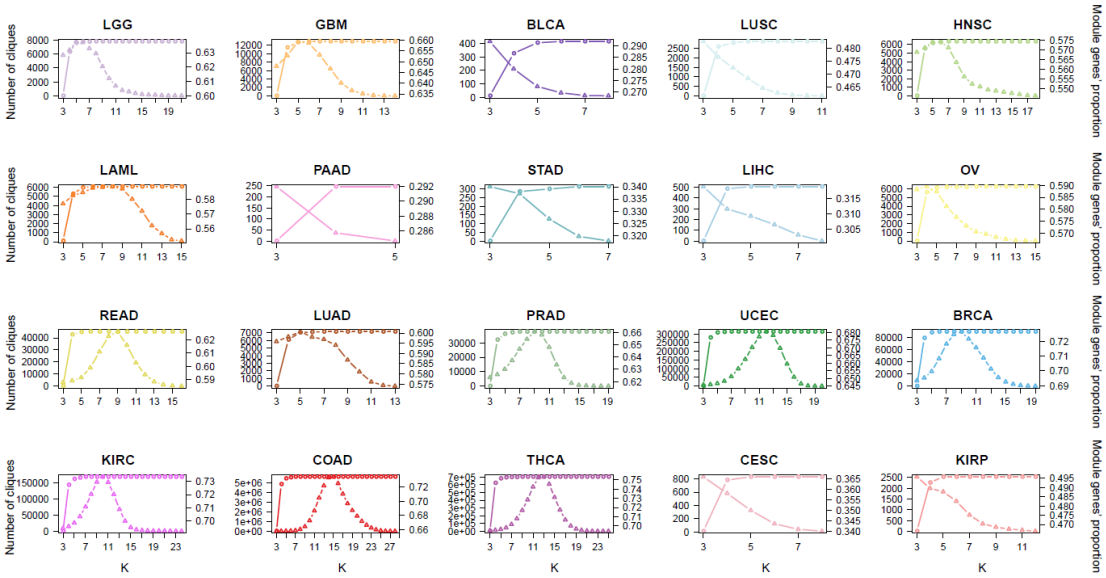

Figure S27. Number of cliques at different k-values and cumulative ratios of ceRNAs in cliques with k-values are not bigger than k. The left y-axis represents number of cliques under different k-values, corresponding to the triangle line. The right y-axis represents cumulative ratios of ceRNAs in cliques, corresponding to the circle line.

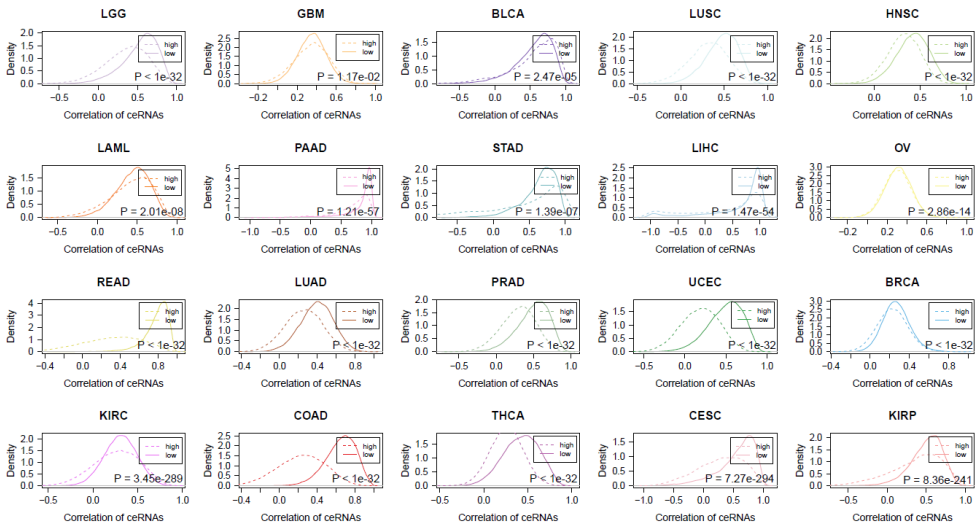

Figure S28. ceRNAs were strongly coexpressed in Dicer- and Drosha-low expressed groups. The solid lines represent the distribution of correlation coefficient of ceRNA pairs in the top 30% samples that are low expressed of Dicer and Drosha; and the dash lines represent the distribution of correlation coefficient of ceRNA pairs in the top 30% samples that are high expressed of Dicer and Drosha. The differences of the two distributions were tested by ranksum test.

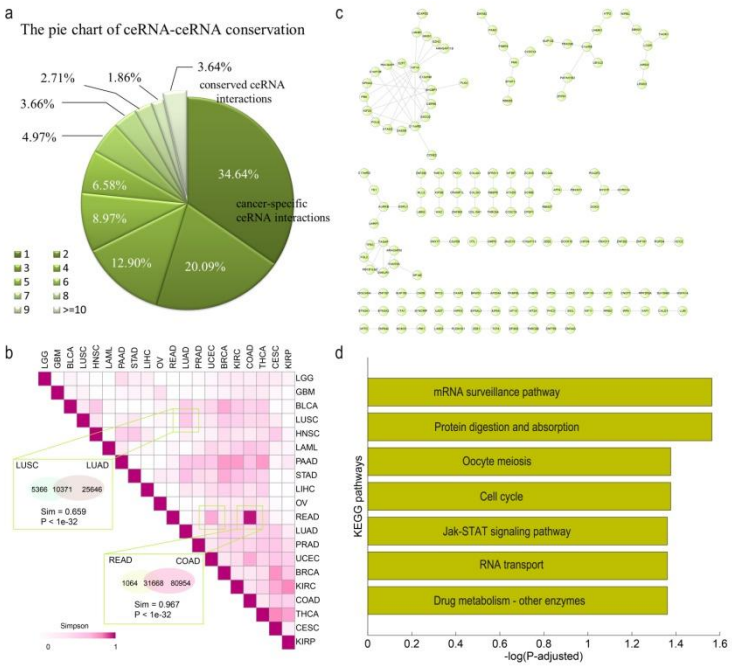

Figure S29. The network level comparision of ceRNA-ceRNA interaction networks across cancers. a, The pie chart shows the proportion of ceRNA interactions presented in different number of cancers. The majority of the ceRNA interactions are cancer specific. **b,** The simpson index matrix shows the similarity between each pair of ceRNA-ceRNA networks. Some pairs of cancers with same origin were specifically shown. **c,** The core neuron ceRNA-ceRNA network that presented in more than 18 cancers. **d,** The KEGG pathways enriched by the genes in the core ceRNA network.

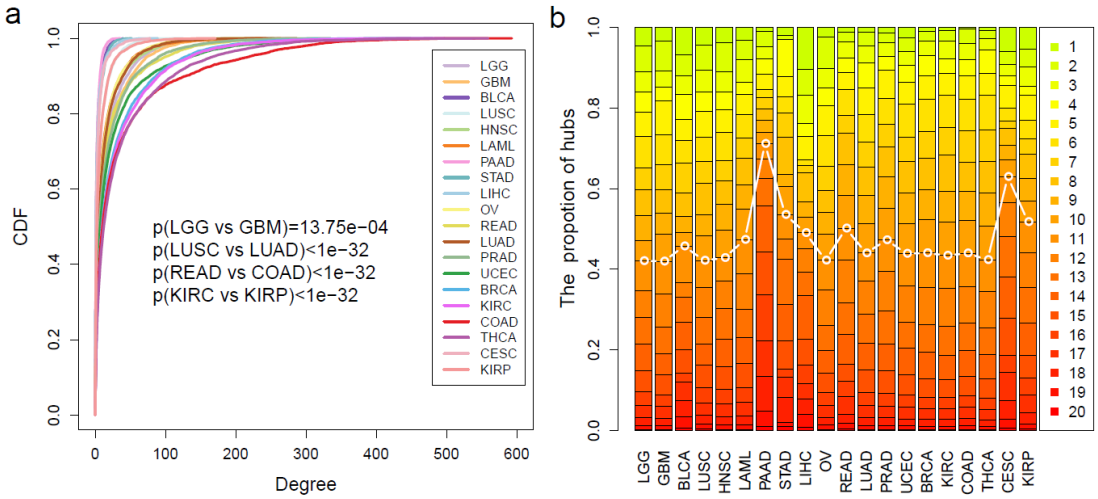

Figure S30. The conserved and rewired network hubs in each cancer type. a, Cumulative distribution functions of the ceRNA degree in each cancer. **b,** The ceRNAs ranked in top 10% were likely to present in top 10% in other cancers.

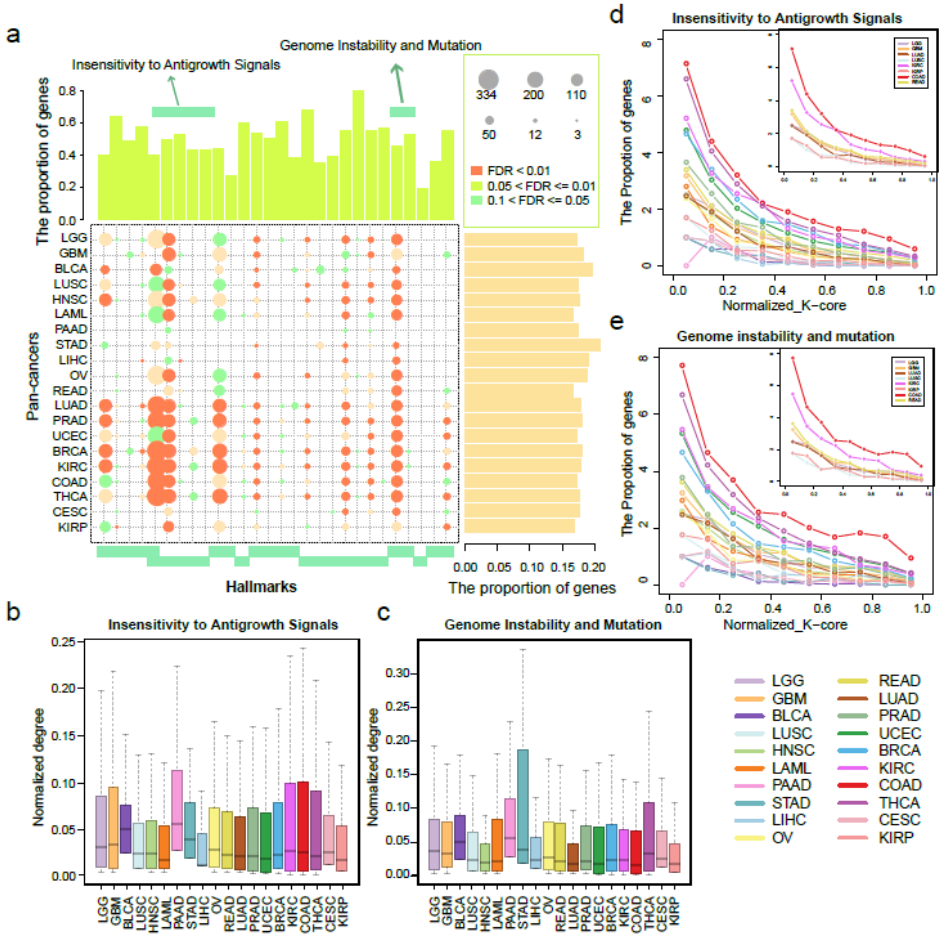

Figure S31. The ceRNA networks control broad cancer associated hallmarks. a, The summary bubble-bar plot show the functional enrichment results of the ceRNA networks across the cancers. **b** and **c**, The normalized degree of ceRNAs annotated in the two hallmarks. **d** and **e**, Relationships between ceRNA layers and frequency of ceRNAs implicated in two hallmarks identified in each layer.

Supplemental Text S3

The structure and functions of ceRNA networks constructed based on multiple molecular datasets.

Construction of the pan-cancer ceRNA networks

Identification of mRNA-mRNA regulation based on integrated analysis

With the increasement of miRNA expression, DNA methylation and DNA copy number available for the same tumors, integration of these information may provide further evidence that the two correlated genes are competitively binding same miRNAs. The multivariate linear model could measure the expression association between a miRNA and a mRNA, that also factors in variation (noise) in mRNA expression induced by changes in DNA copy number and promoter methylation at the mRNA gene locus. In this regression model, the gene expression, changes as a linear function of DNA copy number, DNA methylation and miRNA expression. Then we used the ordinary least square method to obtain an estimate for the coefficient of miRNA, and test the null hypothesis the expression of the miRNA is not associated with change in expression of this gene. The association of miRNA and mRNA were obtained in ten types of cancers (**Table S7**) and a miRNA-mRNA pair was considered as associated if the FDR is under 0.05. And then integrated with the CLIP-seq supported target sites in the main text, we obtained the cancer specific miRNA-mRNA regulations. Next, we performed the same procedure (by considering the shared miRNAs and coexpression) and reconstructed the ceRNA-ceRNA networks in each cancer.

Coregulation of mRNA-related ceRNAs

A hypergeometric test is executed for each possible gene pairs separately. For each given gene pair of gene A and B, we identified the common miRNA regulated them ($A \cap B$). The subset is required to have at least O_{min} miRNAs. And then the probability P for gene A and P is calculated according to

$$P = 1 - F(x \mid N, K, M) = 1 - \sum_{t=0}^{x-1} \frac{\binom{K}{t}\binom{N-K}{M-t}}{\binom{N}{M}}$$

where N is the number of all miRNAs, K and M is the total number of miRNAs regulated gene A and B, x is the common miRNA number between these two genes. Only gene pairs that regulated by at least three common miRNAs were analyzed in our study. All P-values were subject to false discovery rate (FDR) correction. In order to obtain more ceRNA pairs in the networks constructed by integration of multiple molecular datasets. Here, the FDR was relaxed to 0.05.

Coexpression of mRNA-related ceRNAs

Next, integrated with gene expression profiles in all individual cancer types, we identified the ceRNA pairs in specific tumor type. To explore the active ceRNA pairs in individual cancer, we computed the correlation coefficient (R) of each candidate ceRNA pairs identified above. All the candidate gene pairs with R>0 and p-adjusted<0.05 were identified as ceRNA-ceRNA interactions. After assembling all significant ceRNA pairs, we generated the ceRNA network for each cancer type. A node represents a gene, and two nodes are connected if they were coregulated by miRNAs and co-expressed in this cancer.

Table S7. Summary of analyzed TCGA cancer types and data sets.

Cancer	miRNA	mRNA	Methylation	CNA
GBM	380	380	211	380
OV	509	509	509	509
COAD	181	181	181	181
KIRC	368	368	223	368
LUSC	195	195	106	195
BRCA	671	671	415	671
UCEC	332	332	234	332
BLCA	94	94	94	94
HNSC	298	298	298	298

LUAD	347	347	283	347

miRNA-mediated ceRNA interactions in pan-cancer

The multivariate linear model could measure the expression association between a miRNA and a mRNA, that also factors in variation (noise) in mRNA expression induced by changes in DNA copy number and promoter methylation at the mRNA gene locus. The association of miRNA and mRNA were obtained in ten types of cancers and a miRNA-mRNA pair was considered as associated if the FDR is under 0.05. And then we measured the coexpression of mRNAs and reconstructed the ceRNA networks in ten cancers (**Figure S32**). As a result, we found that the ceRNA networks are smaller than those of in the main text. This result suggests that integration of more types of genomic datasets, the noise in the datasets may be filtered. In addition, we found that the majority of these ceRNA networks were also included in our original analysis (Simpson index ranged from 0.25 to 0.53), indicating the results obtained in our main text are robust.

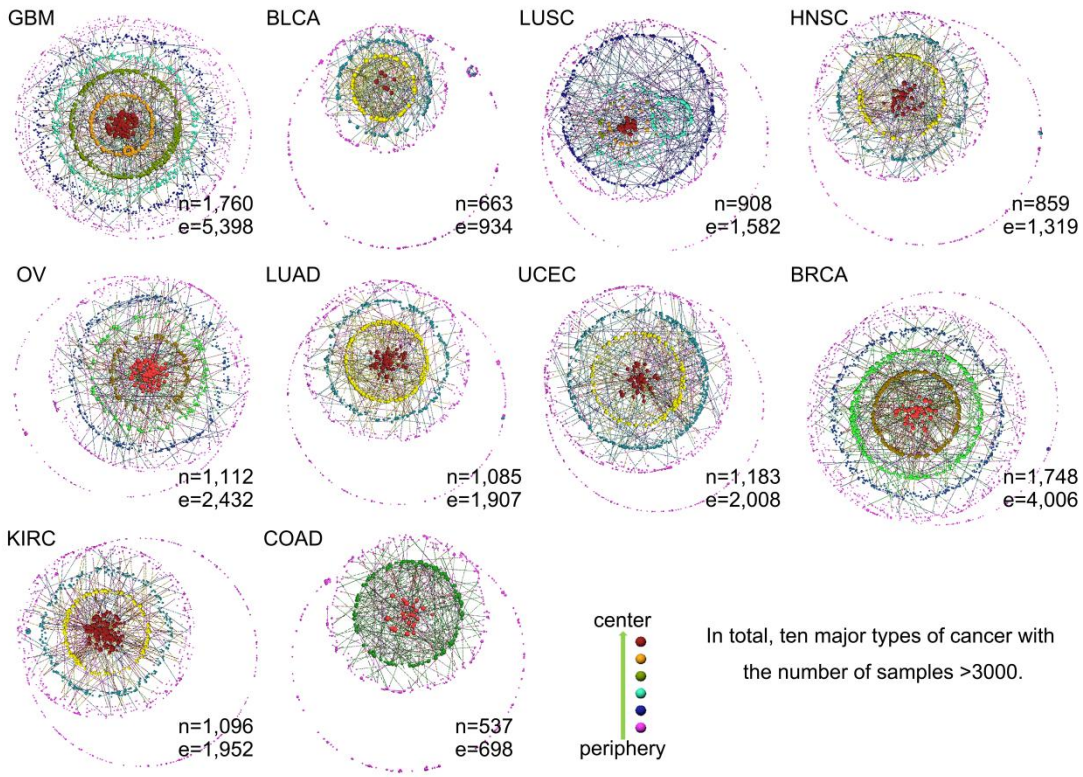

Figure S32. The mRNA-related ceRNA networks in ten types of cancer. Its graphic visualization uses nodes to represent individual ceRNAs and edges to represent miRNA-mediated RNA-RNA interactions. Nodes near the center of the

graph are contained within more tightly regulated, dense subnetworks. The color bands which include nodes with similar connectivity, have a size increases with the distance from the center. The networks are visualized with the Lanet plugin in the network workbench.

Common features of ceRNA interactomes

Analysis of the topological features of the ceRNA networks across cancers, we found that the results obtained in the main text were robust in these newly constructed ceRNA networks. Firstly, the examination of the degree distributions of these ceRNA networks reveals a power law distribution, showing that the ceRNA networks are scale free (**Figure S33**). Secondly, analysis of the ceRNA networks show that highly connected ceRNAs are more coexpressed with their neighbors than others (**Figure S34**). In addition, coexpression of ceRNAs in the network increases with the number of common miRNAs (**Figure S35**). Next, we analyzed the modular structure of the ceRNA networks. All modules in the ceRNA networks are also identified using cFinder. As a result, we found that with an increase in the value of k, there is a sharp decrease in the number of modules. In total, about 25%-48% ceRNAs are involved in at least one module (**Figure S36**). Analysis of the Dicer/Drosha expression, we found that ceRNA pairs were strongly coexpressed in Dicer/Drosha-low expressed groups (**Figure S37**). All these results suggest that the topological features of these ceRNA networks were robust.

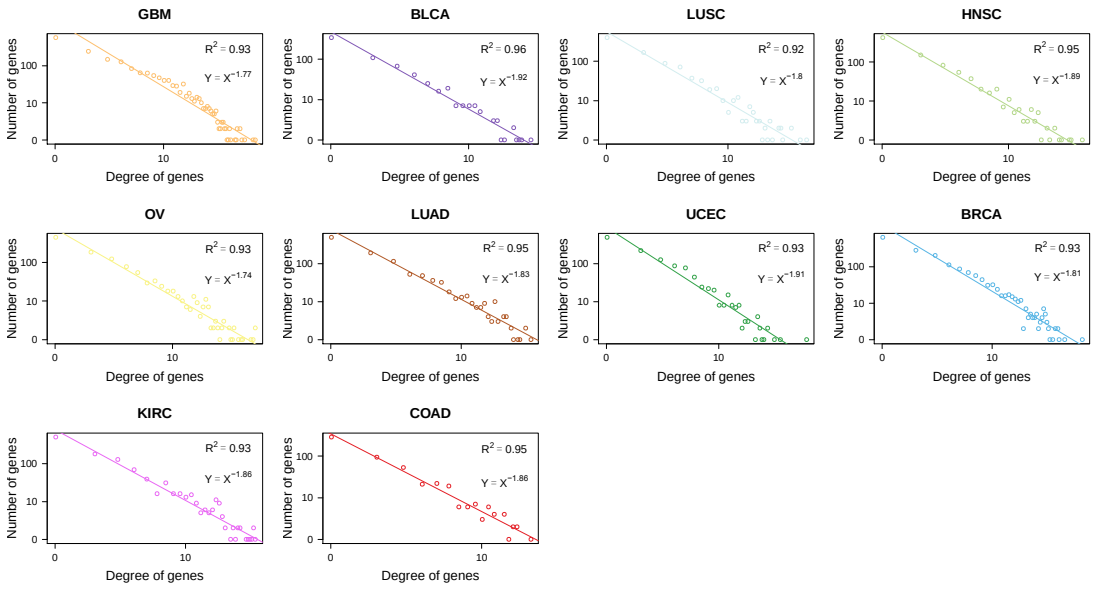

Figure S33. The degree distributions of the ceRNA networks across cancers.

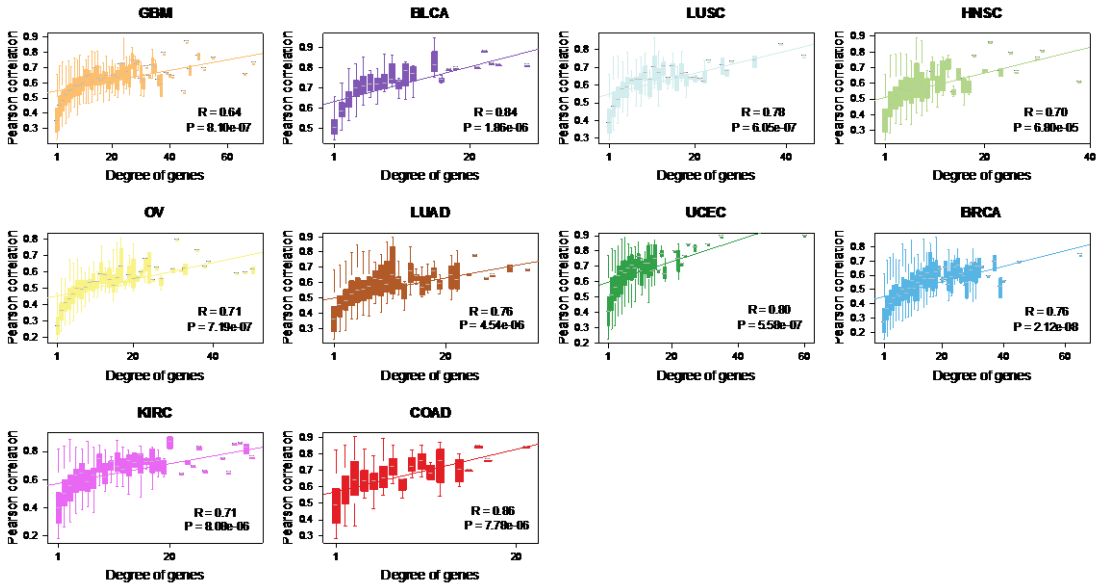

Figure S34. Hub ceRNAs are more coexpressed with their neighbors than others. The correlation between expression of ceRNAs and the total expression of their ceRNA regulators is plotted as a function of the number of its ceRNA regulators; genes at the center of the ceRNA network are regulated by hundreds of ceRNA regulators and are significantly correlated with their total expression.

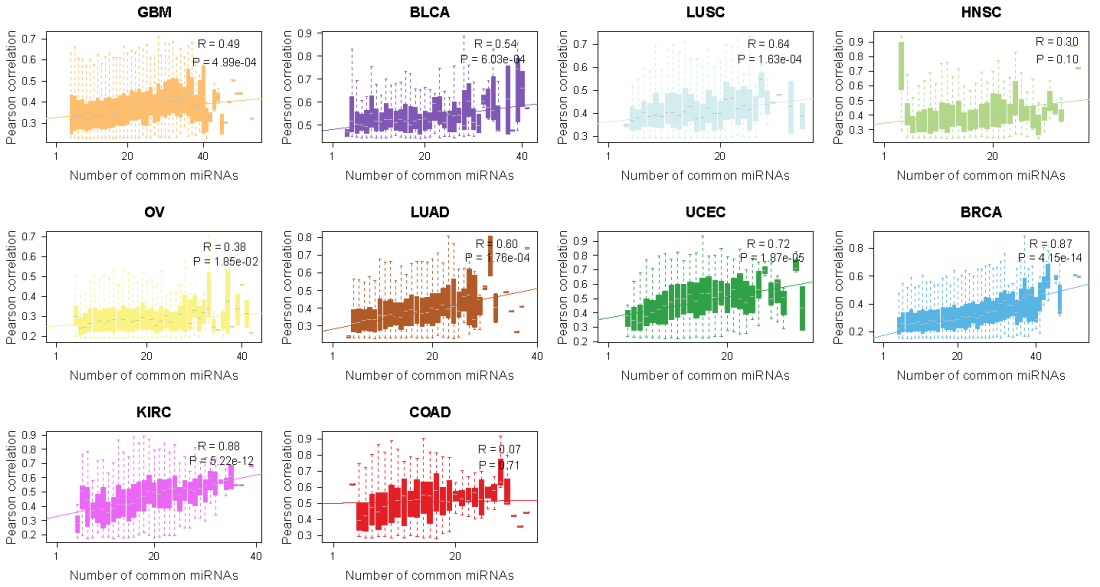

Figure S35. Coexpression of ceRNAs in the network increases with the number of common miRNAs. The ceRNA interactions were grouped by the number of miRNAs they share, and then the correlation coefficient of expressions were shown as boxplot in each group.

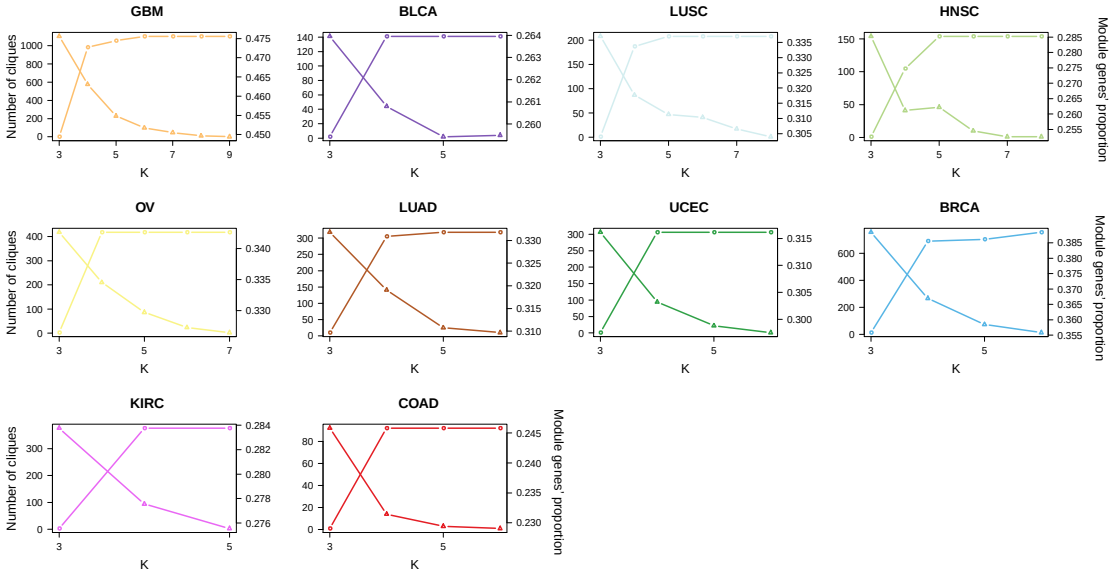

Figure S36. Number of cliques at different k-values and cumulative ratios of ceRNAs in cliques with k-values are not bigger than k. The left y-axis represents number of cliques under different k-values, corresponding to the triangle line. The right y-axis represents cumulative ratios of ceRNAs in cliques, corresponding to the circle line.

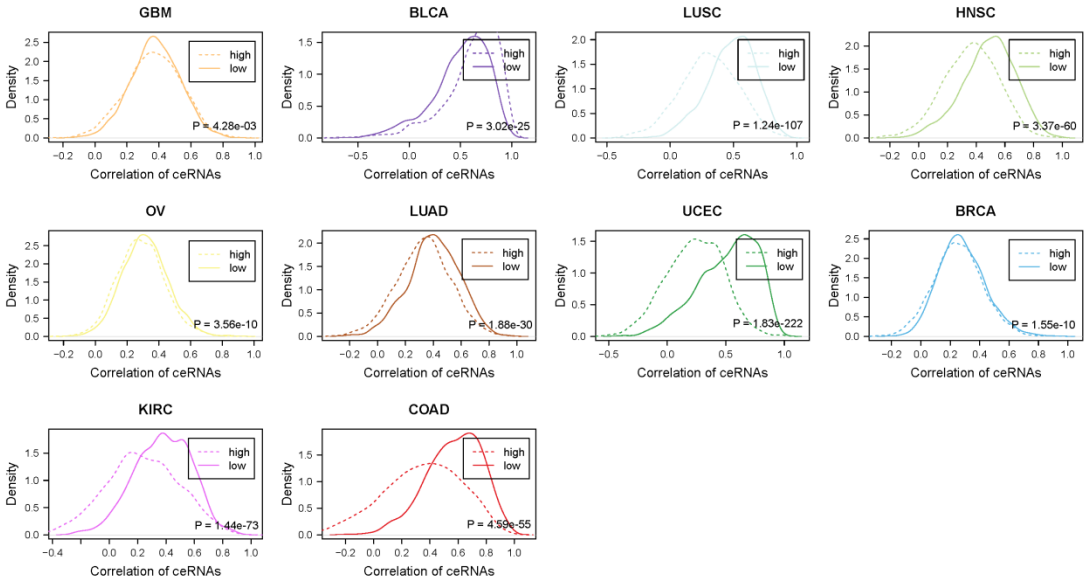

Figure S37. ceRNAs were strongly coexpressed in Dicer- and Drosha-low expressed groups. The solid lines represent the distribution of correlation coefficient of ceRNA pairs in the top 30% samples that are low expressed of Dicer and Drosha; and the dash lines represent the distribution of correlation coefficient of ceRNA pairs in the top 30% samples that are high expressed of Dicer and Drosha. The differences of the two distributions were tested by ranksum test.

Network level analysis across cancers

Viewing the ceRNA network across cancers, our study highlights a marked rewiring

in the ceRNA program between different cancers. We found that only 0.035% ceRNA-ceRNA interactions were conserved in more than five cancers. The low conservation of ceRNA regulations may be explained in part by the cancer-specific expression of genes. Although most of the ceRNA regulations were cancer-specific, the cancers with similar tissue-of-origin also share common ceRNAs (**Figure S38a**). For instance, as expected, LUAD and lung squamous cell carcinoma (LUSC) are two types of lung cancers, we found that the similarity of their ceRNA networks were higher than those with other cancers. Approximately 11.69% ceRNA interactions in LUSC also worked LUAD, which was significantly higher than expected (p<1.73E-3). In addition, we found that the ceRNA-ceRNA interactions observed in more than two cancers form a large component (**Figure S38b**). And functional enrichment analysis indicates that these ceRNAs play key roles in pathways involved in cancers (**Figure S38c**).

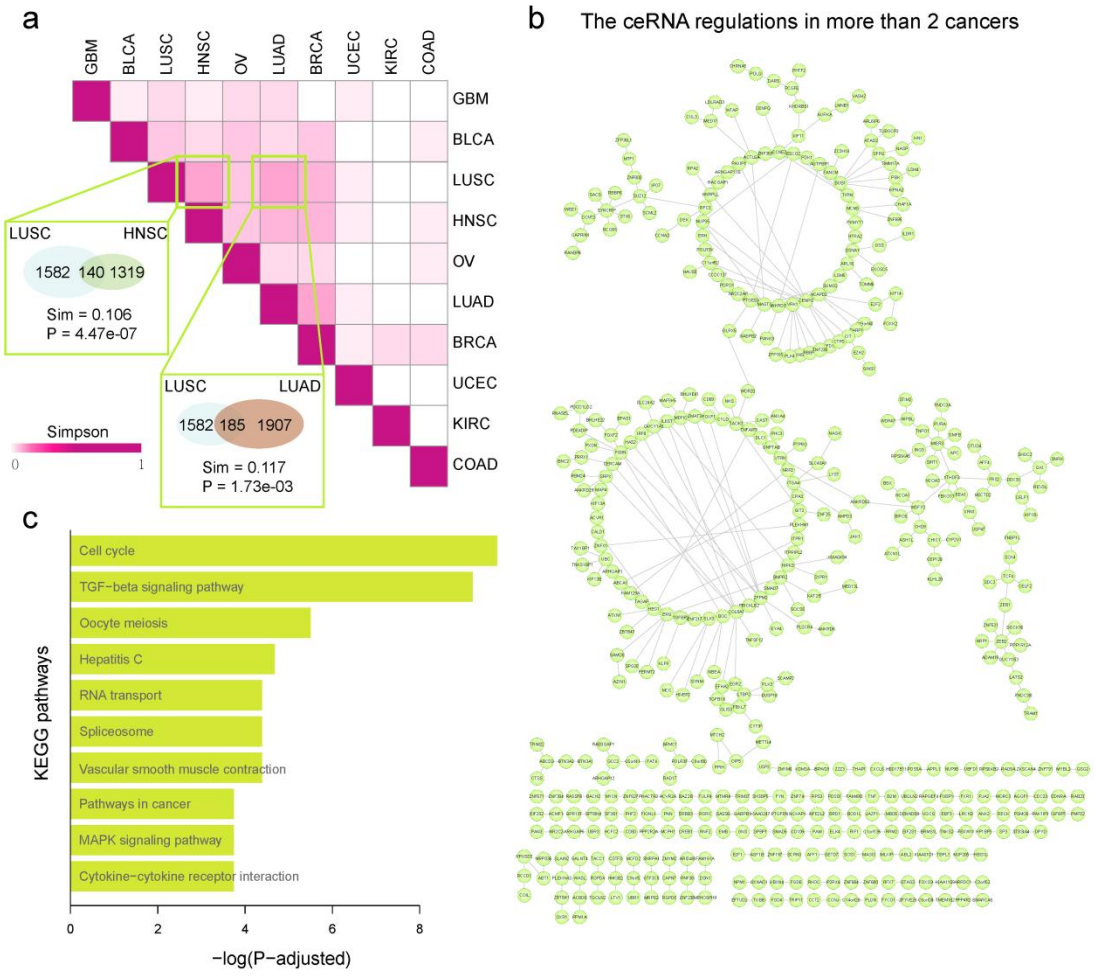

Figure S38. The comparison of ceRNA networks across cancers. a, The network

similarity matrix between two ceRNA networks. **b,** The ceRNA interactions observed in more than two cancers. **c,** The functions of ceRNA enriched.

Differential network analysis

Comparing the degree distribution across cancers, we found that most of the ceRNA networks were characterized by nodes with highly variable degrees, from genes with a few connections to 'hubs' with hundreds of links. Especially, the ceRNA network of glioblastoma multiforme (GBM) presents an increased connectivity with respect to other cancers (**Figure S39a**). Since hub nodes have been found to play important roles in many networks, we also identified hub ceRNAs in each network. Generally, these ceRNA hubs retained their high degree across other cancers (**Figure S39b**). Approximately 20% hubs retained high connectivity in other cancers. In addition, we also identified the common hubs, differential hubs and cancer-specific hubs. When we defined hubs as the top 20% genes with high connectivity, we found that the 62.5% common hubs, 19.2% differential hubs and 22.9% cancer specific hubs were retained in the newly constructed networks. These results suggest that the network structures of the ceRNA networks were robust, these retained hubs may be more important in the development and progression of cancers. Using the more strict roles to filter the candidate ceRNA interactions, we may obtain more confident ceRNAs in cancers.

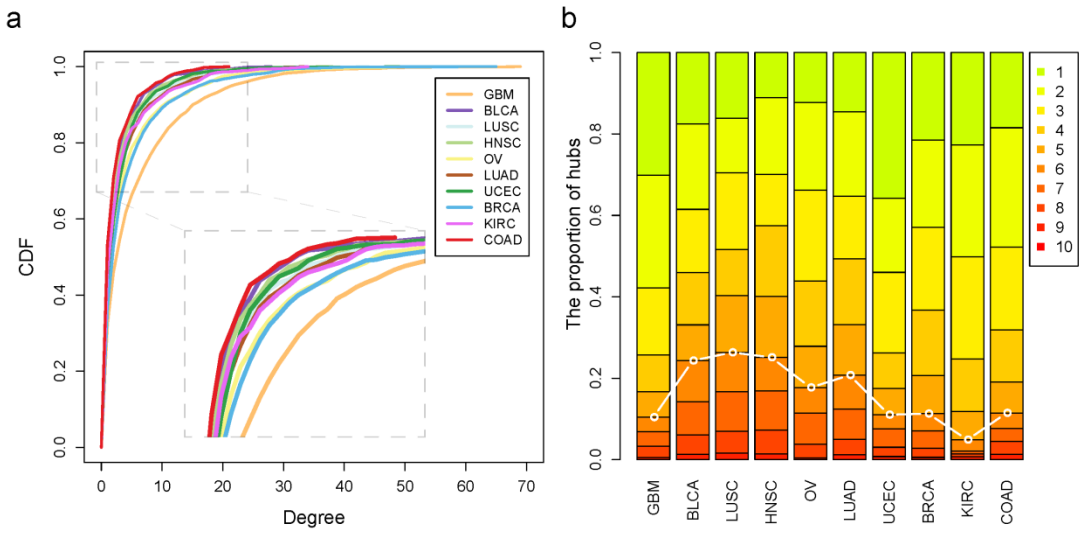

Figure S39. The conserved and rewired network hubs in each cancer type. a, Cumulative distribution functions of the ceRNA degree in each cancer. **b,** The ceRNAs ranked in top 20% were likely to present in top 30-50% in other cancers.

miRNA-mediated ceRNA regulations control broad cancer-related hallmarks

Next, we also performed the functional enrichment analysis of the ceRNA networks across cancers. Functional enrichment analysis reveals that the cancer-related ceRNA networks enriched at least one hallmark of cancers. About 20% hallmark genes were involved in ceRNA regulations, which was significantly larger than randomly chosen genes (**Figure S40a**, right panel, p<1.0E-3). On the other hand, the ceRNA networks cover most genes of the hallmark-related functions (range from 13.54% to 54.54%, **Figure S40a**, top panel). Another interesting observation is that all the ten ceRNA networks are enriched in the function of 'regulation of cell proliferation', highlighting its roles in the development of pan-cancers. Next, we also examined whether the genes enriched in the same hallmarks exhibit different connectivity patterns. The connections number of each ceRNA (node degree) was scaled to a value between 0 and 1 by dividing each node degree by the largest degree in a ceRNAnetwork. We found that the ceRNAs enriched in the same hallmarks show vary degree across cancers (**Figure S40b and S40c**). In addition, by peeling each ceRNA network, we found that the ceRNAs with similar functions localized in different layers of the networks (**S40d and S40e**).

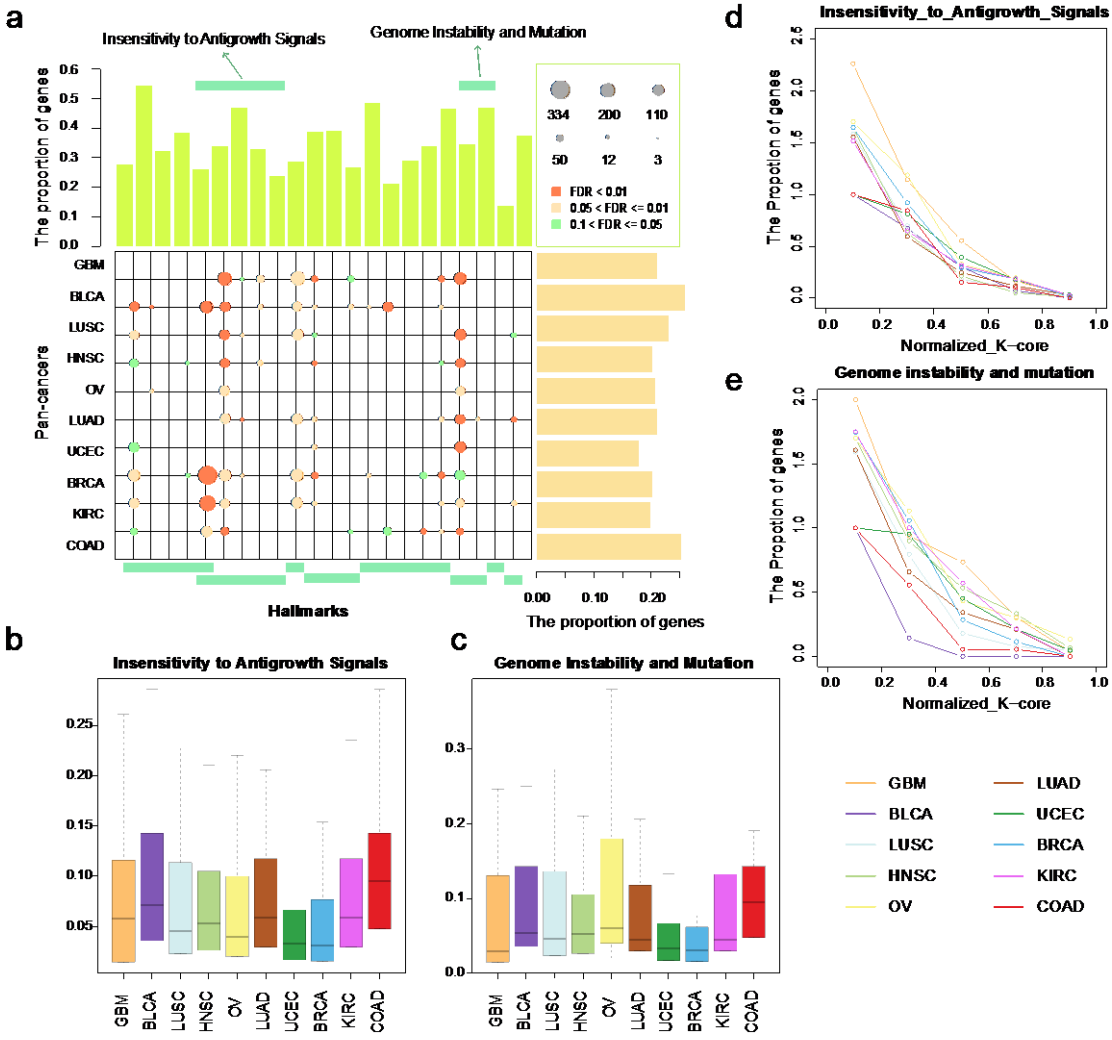

Figure S40. The ceRNA networks control broad cancer associated hallmarks. a, The summary bubble-bar plot show the functional enrichment results of the ceRNA networks across the cancers. **b** and **c**, The normalized degree of ceRNAs annotated in the two hallmarks. **d** and **e**, Relationships between ceRNA layers and frequency of ceRNAs implicated in two hallmarks identified in each layer.

Supplemental references

1. Yang, J.H., Li, J.H., Jiang, S., Zhou, H. and Qu, L.H. (2013) ChIPBase: a database for decoding the transcriptional regulation of long non-coding RNA and microRNA genes from ChIP-Seq data. *Nucleic acids research*, **41**, D177-187.
2. Makinen, V.P., Civelek, M., Meng, Q., Zhang, B., Zhu, J., Levian, C., Huan, T., Segre, A.V., Ghosh, S., Vivar, J. *et al.* (2014) Integrative genomics reveals novel molecular pathways and gene networks for coronary artery disease. *PLoS genetics*, **10**, e1004502.
3. Sumazin, P., Yang, X., Chiu, H.S., Chung, W.J., Iyer, A., Llobet-Navas, D., Rajbhandari, P., Bansal, M., Guarnieri, P., Silva, J. *et al.* (2011) An extensive microRNA-mediated network of RNA-RNA interactions regulates established oncogenic pathways in glioblastoma. *Cell*, **147**, 370-381.
4. Tay, Y., Kats, L., Salmena, L., Weiss, D., Tan, S.M., Ala, U., Karreth, F., Poliseno, L., Provero, P., Di Cunto, F. *et al.* (2011) Coding-independent regulation of the tumor suppressor PTEN by

competing endogenous mRNAs. *Cell*, **147**, 344-357.

5. Karreth, F.A., Tay, Y., Perna, D., Ala, U., Tan, S.M., Rust, A.G., DeNicola, G., Webster, K.A., Weiss, D., Perez-Mancera, P.A. *et al.* (2011) In vivo identification of tumor- suppressive PTEN ceRNAs in an oncogenic BRAF-induced mouse model of melanoma. *Cell*, **147**, 382-395.
6. Sarver, A.L. and Subramanian, S. (2012) Competing endogenous RNA database. *Bioinformation*, **8**, 731-733.
7. de Giorgio, A., Krell, J., Harding, V., Stebbing, J. and Castellano, L. (2013) Emerging roles of competing endogenous RNAs in cancer: insights from the regulation of PTEN. *Molecular and cellular biology*, **33**, 3976-3982.
8. Li, J.H., Liu, S., Zhou, H., Qu, L.H. and Yang, J.H. (2014) starBase v2.0: decoding miRNA-ceRNA, miRNA-ncRNA and protein-RNA interaction networks from large-scale CLIP-Seq data. *Nucleic acids research*, **42**, D92-97.
9. Flores, M., Chen, Y. and Huang, Y. (2014) TraceRNA: a web application for competing endogenous RNA exploration. *Circulation. Cardiovascular genetics*, **7**, 548-557.
10. Poliseno, L. and Pandolfi, P.P. (2015) PTEN ceRNA networks in human cancer. *Methods*, **77-78**, 41-50.
11. Wang, P., Ning, S., Zhang, Y., Li, R., Ye, J., Zhao, Z., Zhi, H., Wang, T., Guo, Z. and Li, X. (2015) Identification of lncRNA-associated competing triplets reveals global patterns and prognostic markers for cancer. *Nucleic acids research*, **43**, 3478-3489.
12. Bandyopadhyay, S. and Mitra, R. (2009) TargetMiner: microRNA target prediction with systematic identification of tissue-specific negative examples. *Bioinformatics*, **25**, 2625-2631.