

RESEARCH ARTICLE

Dynamical network analysis reveals key microRNAs in progressive stages of lung cancer

Chao Kong^{1,2,3}, Yu-Xiang Yao³, Zhi-Tong Bing^{4,5,6}, Bing-Hui Guo⁷, Liang Huang³, Zi-Gang Huang^{1,2*}, Ying-Cheng Lai^{8,9}

1 The Key Laboratory of Biomedical Information Engineering of Ministry of Education, Institute of Health and Rehabilitation Science, School of Life Science and Technology, Xi'an Jiaotong University, The Key Laboratory of Neuro-informatics & Rehabilitation Engineering of Ministry of Civil Affairs, Xi'an, Shaanxi, P. R. China, **2** National Engineering Research Center for Healthcare Devices, Guangzhou, Guangdong, P. R. China, **3** Institute of Computational Physics and Complex Systems, School of Physical Science and Technology, Lanzhou University, Lanzhou, China, **4** Evidence Based Medicine Center, School of Basic Medical Science of Lanzhou University, Lanzhou, China, **5** Key Laboratory of Evidence Based Medicine and Knowledge Translation of Gansu Province, Lanzhou, China, **6** Department of Computational Physics, Institute of Modern Physics, Chinese Academy of Sciences, Lanzhou, China, **7** Beijing Advanced Innovation Center for Big Data and Brain Computing, LMIB and School of Mathematics and System Sciences, Beihang University, Beijing, China, **8** School of Electrical, Computer and Energy Engineering, Arizona State University, Tempe, Arizona, United States of America, **9** Department of Physics, Arizona State University, Tempe, Arizona, United States of America

* huangzg@xjtu.edu.cn



OPEN ACCESS

Citation: Kong C, Yao Y-X, Bing Z-T, Guo B-H, Huang L, Huang Z-G, et al. (2020) Dynamical network analysis reveals key microRNAs in progressive stages of lung cancer. *PLoS Comput Biol* 16(5): e1007793. <https://doi.org/10.1371/journal.pcbi.1007793>

Editor: Ilya Ioshikhes, University of Ottawa, CANADA

Received: December 24, 2019

Accepted: March 17, 2020

Published: May 19, 2020

Peer Review History: PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: <https://doi.org/10.1371/journal.pcbi.1007793>

Copyright: © 2020 Kong et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the manuscript and its Supporting Information files and in the Zenodo repository (<https://doi.org/10.5281/zenodo.3753914>).

Abstract

Non-coding RNAs are fundamental to the competing endogenous RNA (CeRNA) hypothesis in oncology. Previous work focused on static CeRNA networks. We construct and analyze CeRNA networks for four sequential stages of lung adenocarcinoma (LUAD) based on multi-omics data of long non-coding RNAs (lncRNAs), microRNAs and mRNAs. We find that the networks possess a two-level bipartite structure: common competing endogenous network (CCEN) composed of an invariant set of microRNAs over all the stages and stage-dependent, unique competing endogenous networks (UCENs). A systematic enrichment analysis of the pathways of the mRNAs in CCEN reveals that they are strongly associated with cancer development. We also find that the microRNA-linked mRNAs from UCENs have a higher enrichment efficiency. A key finding is six microRNAs from CCEN that impact patient survival at all stages, and four microRNAs that affect the survival from a specific stage. The ten microRNAs can then serve as potential biomarkers and prognostic tools for LUAD.

Author summary

Lung cancer is the leading cause of cancer-related human deaths worldwide. Lung adenocarcinoma is one of the most common subtypes, and has more pronounced genomic variations than other lung cancer subtypes. A milestone discovery in cancer research is the roles played by non-coding RNAs which have been identified as the oncogenic drivers

Funding: ZGH acknowledges supports from NNSF of China under Grants (Nos. 11975178, and 61431012), Natural Science Basic Research Plan in Shaanxi Province of China (Program No. 2020JM-058), Fundamental Research Funds for the Central Universities (sxzd022020012), and support of K. C. Wong Education Foundation. LH acknowledges supports from NNSF of China under Grants Nos. 11775101, and 11422541. YCL would like to acknowledge support from the Vannevar Bush Faculty Fellowship program sponsored by the Basic Research Office of the Assistant Secretary of Defense for Research and Engineering and funded by the Office of Naval Research through Grant No. N00014-16-1-2828. BHG acknowledges support from Artificial Intelligence Project (2018AAA0102301). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

and tumor suppressors. In cancer development, non-coding RNAs form an inseparable unity of RNA-level regulating networks in the intracellular environment, and the dynamical interplay and competition among different types of RNAs are playing a pivotal role. We have developed a quantitative approach to reconstructing the mutual regulation networks of RNAs for the progressive stages of lung adenocarcinoma at the post-transcriptional level. Our analysis revealed the emergence of two characteristically distinct types of networks that possess a two-level bipartite structure, and we uncovered a number of key genes that affect or even determine the survival of patients at each stage. Our work establishes a more comprehensive gene-data analysis framework than previous ones, not only providing a tool to probe more deeply into the mechanisms of cancer evolution than previously possible but also having the potential to lead to more effective biomarkers and drug targets for lung cancer.

Introduction

Lung cancer is the leading cause of cancer-related human deaths worldwide [1]. Approximately 85% of the lung cancer cases can be classified as non small-cell lung cancer (NSCLC), among which lung adenocarcinoma (LUAD) is one of the most common subtypes [2]. The mechanisms behind cancer evolution are extremely complex, impeding accurate and reliable prognosis as well as effective treatment [3]. A standard existing approach to monitoring tumor progress and detecting/ascertaining the underlying mechanism is mRNA transcriptomics, which has led to a large number of significant prognostic biomarkers and therapeutic targets [4–6]. When combined with other omics information, such as microRNA, methylation and epigenetic data, mRNA transcriptomics has provided valuable insights into the mechanisms of lung cancers [7–10].

A milestone discovery in cancer research is the roles played by non-coding RNAs (ncRNAs). In the general developmental and disease contexts, at the mRNA level, ncRNAs have been found in key regulators of physiological functions [11–15]. Particularly relevant to cancer, ncRNAs have been identified as the oncogenic drivers and tumor suppressors in major cancer types [16]. The discovery of the pivotal role of ncRNAs in cancer has led to the paradigmatic, competing endogenous RNA (CeRNA) hypothesis [17–19]: in cancer emergence and development, microRNAs, mRNAs, and ncRNAs form an inseparable unity of RNA-level regulating network in the intracellular environment, collectively known as the CeRNA network. For example, it has been known [8, 18] that RNA-induced silencing complex (RISC) can be produced through binding of microRNA response elements (MREs) at RNAs 3'UTR, causing inactivation of the target mRNA, but this mechanism is also present in non-coding RNAs, where the target RNA usually has multiple MREs. There has been increasing evidence for the fundamental roles played by CeRNAs in biological systems [16, 18] and, as a result, studying cancer-related CeRNA networks based on RNA-sequence expression data has gained momentum. For example, dysregulated CeRNA-CeRNA interactions in the CeRNA networks of LUAD were analyzed [20, 21], suggesting that the gain or loss of CeRNAs can be used in functional analysis and as potential diagnostic biomarkers. The difference in the expression profiles between early and late stages in CeRNA networks of LUAD was noticed and some LUAD specific, long non-coding RNAs (lncRNAs) with their functional enrichment and clinical features were uncovered [22]. The possible roles played by lncRNAs through CeRNA networks in other types of cancer such as liver cancer and papillary thyroid cancer were also studied [5].

In existing studies of CeRNA networks, a commonly practiced methodology is to identify some differentially expressed lncRNAs, mRNAs and microRNAs based on absolute fold changes and values of the false positive ratio. Consequently, information used to construct the underlying CeRNA networks and to reveal the network functions with clinical implications is directly from data bases without any dynamical ingredient. The resulting CeRNA networks are thus simply a combination of different kinds of RNAs, whereas the dynamical interplay and competition among different types of RNAs were completely ignored. To remedy this deficiency has motivated our work.

We hold the belief that cancer development and evolution are fundamentally a dynamic process. Manifested in the underlying CeRNA networks, it is unlikely that the intrinsic interactions, interplay and the network structure remain static during cancer development. To better understand cancer and to identify more effective biomarkers, the dynamical aspects of the CeRNA networks must be taken into account. This is in line with the field of network physiology [23–27], where transitions and reorganization occur in the networks of physiological/organ systems in the human body on larger spatial and temporal scales, which can be constructed using readily accessible continuous time series. In the CeRNA network analysis, it is infeasible to collect RNA data in continuous time. Nonetheless, with the presently available gene data, we are able to incorporate the time axis into the analysis but only for a limited set of time intervals. In particular, we focus on the four stages of LUAD progression to construct, based on both RNA expression and clinical data of LUAD from the cancer genome atlas (TCGA), the lncRNA-microRNA-mRNA CeRNA network for each stage. For any given stage, our quantitative approach to reconstructing the CeRNA network consists of analyzing differentially expressed RNAs, matching microRNA targets by base complementary pairing, and selecting the negative correlation by invoking the CeRNA hypothesis. We then calculate the fold changes and the average expression level of RNAs in the CeRNA networks for the four stages and carry out Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analyses to validate the results. Our analysis reveals the emergence of two characteristically distinct types of networks that play an important role in the progression of LUAD from stage I to stage IV: one is common to all four stages, which we name as the common competing endogenous RNA network (CCEN), and another is unique for each stage, which we call as the unique competing endogenous RNA networks (UCENs). Analyzing the properties of CCEN and UCENs, we uncover a number of key genes that affect or even determine the survival of patients at each stage of LUAD: six microRNAs from the CCEN which affect the survival of LUAD patients at all stages, and four other microRNAs that influence the survival at each specific stage. Our work establishes a more comprehensive gene-data analysis framework than previous ones, not only providing a tool to probe more deeply into the mechanisms of cancer evolution than previously possible but also having the potential to lead to more effective biomarkers and drug targets for LUAD as well as other types of cancer.

Results

Reconstructed lncRNA-microRNA-mRNA CeRNA networks and doubly bipartite structure

As described in **Materials and Methods**, to reconstruct the lncRNA-microRNA-mRNA CeRNA networks, we first obtain the differentially expressed RNAs by comparing the gene expression values in the LUAD samples with those from the normal samples of lncRNAs, microRNAs and mRNAs. We then apply the principle of base complementary pairing to match the microRNAs with lncRNAs and mRNAs, all differentially expressed, leading to values of the correlation between microRNA-lncRNA and microRNA-mRNA nodes. Retaining

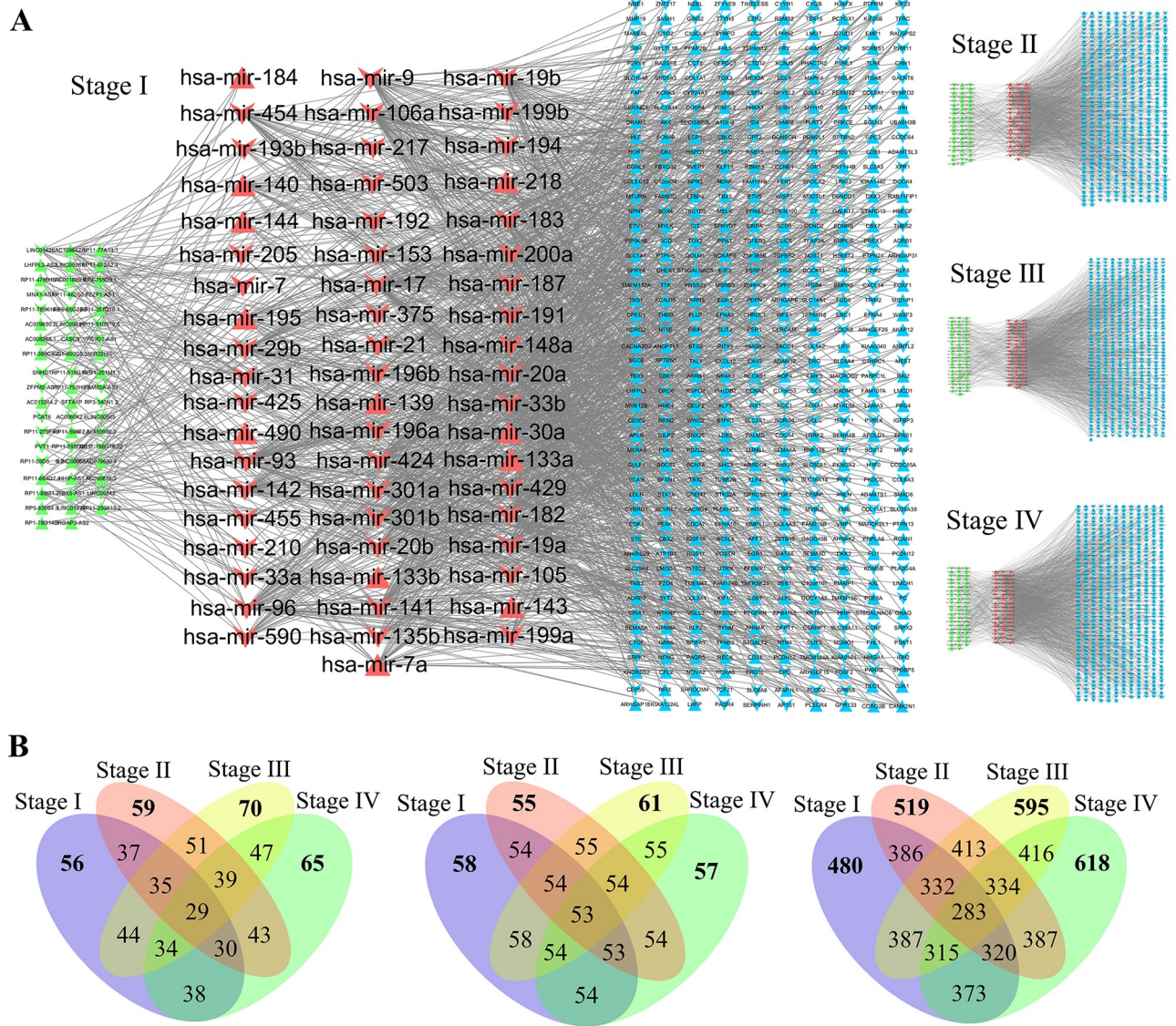


Fig 1. Doubly bipartite or mutualistic structure of the lncRNA-microRNA-mRNA CeRNA networks. (A) Reconstructed lncRNA-microRNA-mRNA CeRNA networks for the four stages of LUAD, where the green, red, and blue nodes indicate lncRNAs, microRNAs, and mRNAs, respectively. The edges represent the mutual regulation between microRNAs and lncRNAs or between microRNAs and mRNAs, which are from the data bases of microRNA targets. At a larger scale where the clusters of lncRNA, microRNA, and mRNA nodes are viewed as three “supernodes,” there is mutualism because the interaction between the lncRNA and mRNA supernodes is through the microRNA supernode. Mutualism also arises at the smaller scale of individual RNA nodes. Especially, the lncRNAs and microRNAs (or the mRNAs and microRNAs) constitute a bipartite network where the interactions between any two nodes of the same color must be through a node of a different color. (B) Venn diagrams illustrating the numbers of nodes in the lncRNA (left), microRNA (middle) and mRNA (right) CeRNA networks of the four stages of LUAD, reflecting the dynamical evolution of the networks.

<https://doi.org/10.1371/journal.pcbi.1007793.g001>

only the negatively correlated pairs of microRNA-lncRNAs and microRNA-mRNAs, we obtain the final CeRNA networks. (S1 Table presents more details of the network reconstruction process).

Fig 1 presents examples of the reconstructed lncRNA-microRNA-mRNA CeRNA networks for the four stages of LUAD. A common feature for the networks in all four states is that direct connections exist only between microRNAs (red nodes) and lncRNAs (green nodes) or between the microRNAs and mRNAs (blue nodes). The connections between lncRNAs and

mRNAs are thus indirect: they occur through the microRNAs. Regarding the lncRNAs, mRNAs, and microRNAs as three entities (or “supernodes”), we see that the connection between the first two supernodes is through the third one, which is characteristic of mutualistic or bipartite type of interactions. The mutualism occurs at a finer scale because the subnetwork of microRNAs-lncRNAs (or that of microRNAs-mRNAs) also has a bipartite structure: there are no direct links among the microRNAs or among the lncRNAs, i.e., any connection between two nodes of the same color must be indirect and through a node of a different color. Such bipartite network structure arises commonly in ecology, e.g., the pollinator-plant mutualistic networks [28–36]. The striking feature here is that the lncRNA-microRNA-mRNA CeRNA networks for all four stages of LUAD possess a doubly bipartite structure: mutualism occurs at two distinct scales.

Changes in fold-change value and expression

Compared with RNAs from the direct differential expression analysis, the differentially expressed RNAs of the final CeRNA networks obtained through matching the differentially expressed microRNA targets and retaining only those RNA pairs with negative correlation have higher values of the expression of microRNAs and fold change, as shown in Fig 2. Our

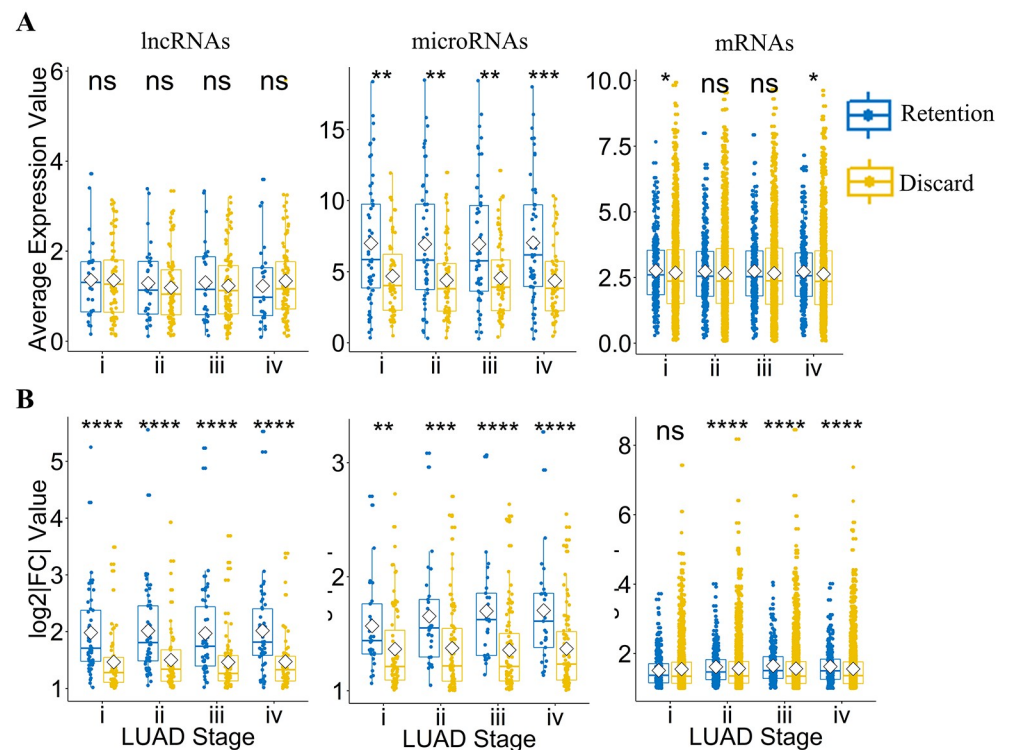


Fig 2. Gene expression and fold-change values of lncRNA, microRNA and mRNA nodes in the reconstructed CeRNA networks of the four stages of LUAD. The two-sample Wilcoxon test (also known as “Mann-Whitney” test) [37, 38] is applied to vectors of the RNA expression data, with the following notations for statistical significance: ns— $p > 0.05$; *— $p \leq 0.05$; **— $p \leq 0.01$; ***— $p \leq 0.001$; ****— $p \leq 0.0001$. The blue group “Retention” represents RNAs selected through analyzing differentially expressed (DE) RNAs, matching microRNA targets, and selecting the negative correlation in the CeRNA networks. For comparison, the yellow group “Discard” describes RNAs processed only through the differential expression analysis, which are not used in the construction of the CeRNA networks. The hypothesis tests between the “Discard” and “Retention” groups are of the Wilcoxon type. (A) Gene expression values of lncRNAs, microRNAs and mRNAs in the reconstructed CeRNA networks of the four stages of LUAD. (B) The corresponding fold-change values of lncRNAs, microRNAs and mRNAs.

<https://doi.org/10.1371/journal.pcbi.1007793.g002>

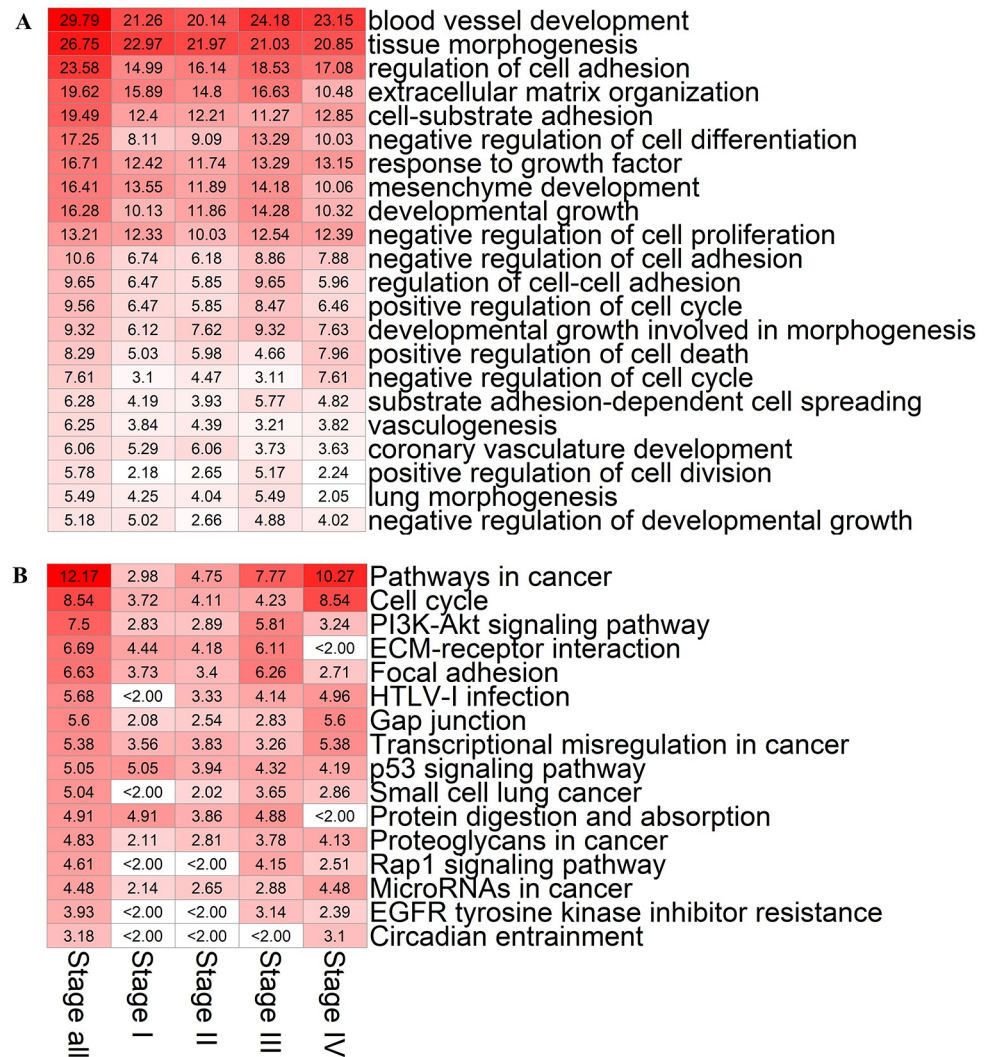


Fig 3. Gene enrichment analysis of mRNAs in the CeRNA networks of four stages of LUAD. The first column in the heat map indicates the results of enrichment analysis. The second to fifth columns represent the results of analysis from the first to the fourth LUAD stage, respectively. The values of the heat map are those of $-\log_{10} P$ from the enrichment analysis. The hypergeometric distribution test is used to calculate the P-value. (A) Results of gene ontology enrichment analysis of various biological processes. (B) Results of KEGG enrichment analysis.

<https://doi.org/10.1371/journal.pcbi.1007793.g003>

construction of the CeRNA networks thus result in more deviated RNAs relative to normal sample expression of LUAD.

Validation of network reconstruction: Enrichment analysis of mRNAs in CeRNA networks

Are the lncRNA-microRNA-mRNA CeRNA networks so constructed meaningful? To address this question, we perform gene ontology and KEGG pathway gene enrichment analysis of mRNAs in the CeRNA networks at the four stages of LUAD, with results in Fig 3. We find that the mRNAs in the CeRNA networks at all four stages are strongly correlated with blood vessel development ($-\log_{10} P > 20$), tissue morphogenesis ($-\log_{10} P > 20$) and regulation of cell adhesion ($-\log_{10} P > 15$), where P is the P-value (Materials and methods). Strong correlation

($-\log_{10} P > 10$) also exists between the mRNAs and factors such as extracellular matrix organization, cell-substrate adhesion, response to growth factor, mesenchyme development, developmental growth and negative regulation of cell proliferation. Remarkably, beyond the “static” information provided by the conventional gene ontology analysis of the four stages of LUAD, our CeRNA networks give rise to a dynamic scenario for tumor progression as the TNM stage deteriorates (see [Materials and methods](#) for the meaning of TNM). For example, mRNAs are more correlated with coronary vasculature development and less correlated with substance-dependent cell spreading in the early than the late two stages. In addition, mRNAs are more correlated with positive regulation of cell death and negative regulation of cell cycle ($-\log_{10} P > 7$) and less correlated with lung morphogenesis ($-\log_{10} P < 6$) in the late stages than the early three stages of LUAD.

The KEGG pathway enrichment analysis reveals that the mRNAs associated with pathway in cancer ($-\log_{10} P = 2.98, 4.75, 7.77, \text{ and } 10.2$ for the four stages, respectively) and cell cycle ($-\log_{10} P = 3.72, 4.11, 4.23, \text{ and } 8.54$ for the four stages, respectively) have stronger correlation in early than late TNM stages of LUAD. Additionally, some cancer-related pathways in the late stages are more significant than in the early stages, such as PI3K-Akt signaling pathway, focal adhesion, gap junction, transcriptional misregulation in cancer, proteoglycans in cancer, Rap1 signaling pathway, microRNAs in cancer, EGFR tyrosine kinase inhibitor resistance, and circadian entrainment. On the contrary, ECM-receptor interaction as well as protein digestion and absorption are less relevant to the LUAD late stages than to the early stages.

Taken together, the enrichment analysis of mRNAs reveals that our reconstructed lncRNA-microRNA-mRNA CeRNA networks exhibit a close correspondence to the development of LUAD and deterioration of physiological indicators from stage I to stage IV, validating our reconstruction method.

CCEN and UCEN analysis

To better understand the relationship between CeRNA networks and TNM stages of LUAD, we extract the CCEN and UCENs from the lncRNA-microRNA-mRNA CeRNA networks, as shown in [Fig 4](#).

[Fig 5A](#) shows the results of gene ontology enrichment analysis. We see that the mRNAs of the UCENs are correlated with blood vessel development, tissue morphogenesis, regulation of cell adhesion, extracellular matrix organization, cell-substrate adhesion and regulation of growth ($-\log_{10} P > 5$). Consistent with the results of enrichment analysis of the mRNAs in the CeRNA networks ([Fig 3](#)), the UCENs can also distinguish the biological characteristics between early and later stages of the cancer. In particular, the UCENs from the late stages are more correlated with regulation of growth, negative regulation of cell adhesion, positive regulation of cell death, substrate adhesion-dependent cell spreading, regulation of cell division and vascular process in circulatory system than those in the early stage. However, extracellular matrix organization and mesenchyme development in the fourth stages do not exhibit significant changes.

From the KEGG pathway enrichment analysis ([Fig 5B](#)), we find that the CCEN network is relevant to pathways in cancer, small cell lung cancer, PI3K-Akt signaling pathway, ECM-receptor interaction, focal adhesion and cell cycle, while all these pathways belonging to the later two stages show a stronger correlation than those in the early two stages.

Enrichment efficiency of microRNA-linked mRNAs in UCENs

Our reconstructed CeRNA networks, UCEN, and CCEN have 868, 585, and 283 mRNAs altogether, respectively. In comparison, in the UCEN there are 201 microRNA-linked mRNAs.

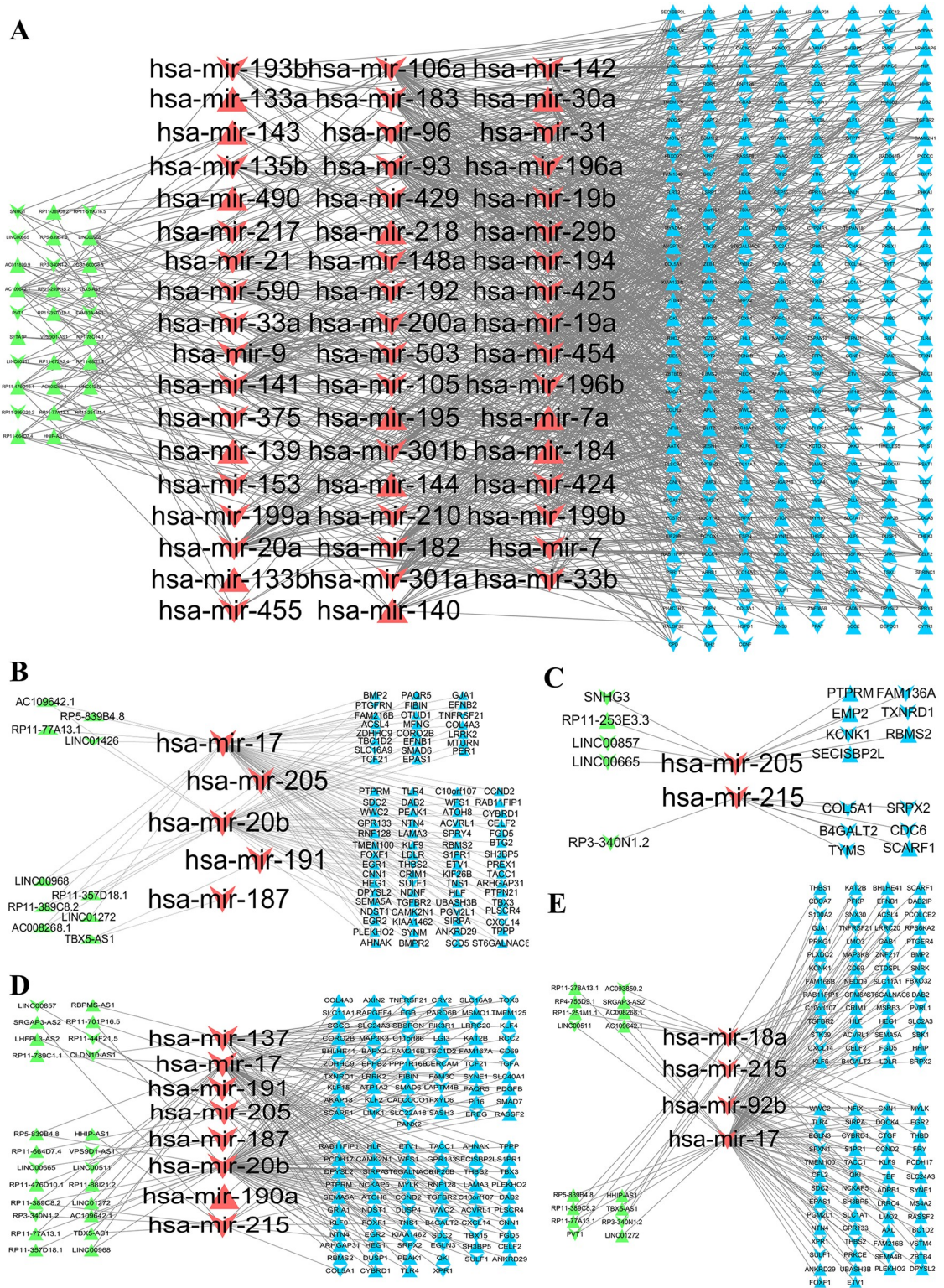


Fig 4. Representative CCEN and UCENs. (A) CCEN extracted from the lncRNA-microRNA-mRNA CeRNA networks of the four stages of LUAD. (B-E) UCENs corresponding to the four stages. The green, red, and blue nodes represent lncRNAs, microRNAs, and mRNAs, respectively. The edges represent mutual regulation between microRNA-lncRNA and between microRNA-mRNA from the microRNA target data bases.

<https://doi.org/10.1371/journal.pcbi.1007793.g004>

	All	Common	Stage I	Stage II	Stage III	Stage IV	
A	29.79	14.76	7.16	6.31	10.07	9.23	blood vessel development
	26.75	13.87	9.56	8.65	7.99	7.9	tissue morphogenesis
	23.58	10.05	5.56	6.69	9	7.68	regulation of cell adhesion
	19.62	5.36	11.88	10.37	12.06	5.65	extracellular matrix organization
	19.49	5.52	7.65	7.32	6.27	7.84	cell-substrate adhesion
	18.65	11.11	2.13	2.37	4.72	5.78	positive regulation of locomotion
	17.25	4.28	4.51	5.51	9.95	6.39	negative regulation of cell differentiation
	16.71	9.19	3.98	3.51	5	4.93	response to growth factor
	16.41	6.1	8.16	6.31	8.56	4.56	mesenchyme development
	15.66	4.1	5.93	7	6.9	7.88	regulation of growth
	13.21	8.16	4.81	2.87	5.16	5.07	negative regulation of cell proliferation
	10.71	3.69	3.03	3.98	4.79	3.87	pattern specification process
	10.67	6.61	<2.00	4.56	2.77	2.49	gland development
	10.6	3.96	3.33	2.78	5.46	4.45	negative regulation of cell adhesion
	10.55	4.58	4.01	2.77	4.15	6.42	rhythmic process
	9.56	4.78	2.31	<2.00	4.27	2.47	positive regulation of cell cycle
	8.29	3.7	<2.00	2.89	<2.00	4.87	positive regulation of cell death
	7.88	3.84	<2.00	<2.00	2.88	3.25	regulation of cell cycle process
	7.72	<2.00	3.32	2.17	5.13	2.45	regulation of transmembrane transport
	6.28	3.22	<2.00	<2.00	3	2.12	substrate adhesion-dependent cell spreading
	6.25	<2.00	2.44	3.02	<2.00	2.36	vasculogenesis
	5.85	<2.00	<2.00	<2.00	5.85	2.12	regulation of cell division
	5.84	3.19	<2.00	<2.00	<2.00	2.72	vascular process in circulatory system
B	12.17	3.07	<2.00	2.24	5.35	8.05	Pathways in cancer
	7.5	<2.00	2.68	2.68	6.23	2.97	PI3K-Akt signaling pathway
	7.01	3.96	<2.00	<2.00	<2.00	5.03	Cell cycle
	6.97	<2.00	5.31	4.8	6.97	<2.00	ECM-receptor interaction
	6.63	<2.00	2.89	2.45	5.67	<2.00	Focal adhesion
	5.84	5.84	<2.00	<2.00	<2.00	<2.00	p53 signaling pathway
	5.68	2.47	<2.00	<2.00	2.2	3.02	HTLV-I infection
	5.53	<2.00	<2.00	<2.00	2.24	5.43	Gap junction
	5.04	<2.00	<2.00	<2.00	3.11	2.2	Small cell lung cancer
	4.7	<2.00	<2.00	<2.00	4.7	2.56	Rap1 signaling pathway
	4.55	3.54	<2.00	<2.00	<2.00	2.42	Transcriptional misregulation in cancer
	4.13	<2.00	<2.00	<2.00	4.13	3.09	EGFR tyrosine kinase inhibitor resistance
	4.04	2.38	3.06	<2.00	2.95	<2.00	Protein digestion and absorption
	3.18	<2.00	<2.00	<2.00	<2.00	2.65	Circadian entrainment

Fig 5. Gene enrichment analysis of mRNAs in the CCEN and UCENs associated with the four stages of LUAD. The first column “All” in the heat map presents the results of enrichment analysis in the four LUAD stages, and the column “Common” lists the results of enrichment analysis in the CCEN. The columns “Stage I” to “Stage IV” show the results of mRNAs in the four stages. The value of the heat map is $-\log_{10} P$ from the enrichment analysis. The P-value is calculated using the hypergeometric distribution test. (A) Results from gene ontology enrichment analysis of various biological processes. (B) Results from KEGG enrichment analysis.

<https://doi.org/10.1371/journal.pcbi.1007793.g005>

We find that they offer the best average enrichment efficiency in blood vessel development, regulation of cell adhesion, circulatory system process, tissue morphogenesis and pathways in cancer, as shown in Fig 6. The microRNA-associated CeRNA networks with higher enrichment efficiency thus exhibit stronger correlation with LUAD development than other types of CeRNA networks.

MicroRNAs in CCEN with a significantly influence on patient survival

We perform the Kaplan-Meier analysis [39] and find that there are six microRNAs that have a significant effect on patient survival for all stages of LUAD, as shown in Fig 7. They are hsa-

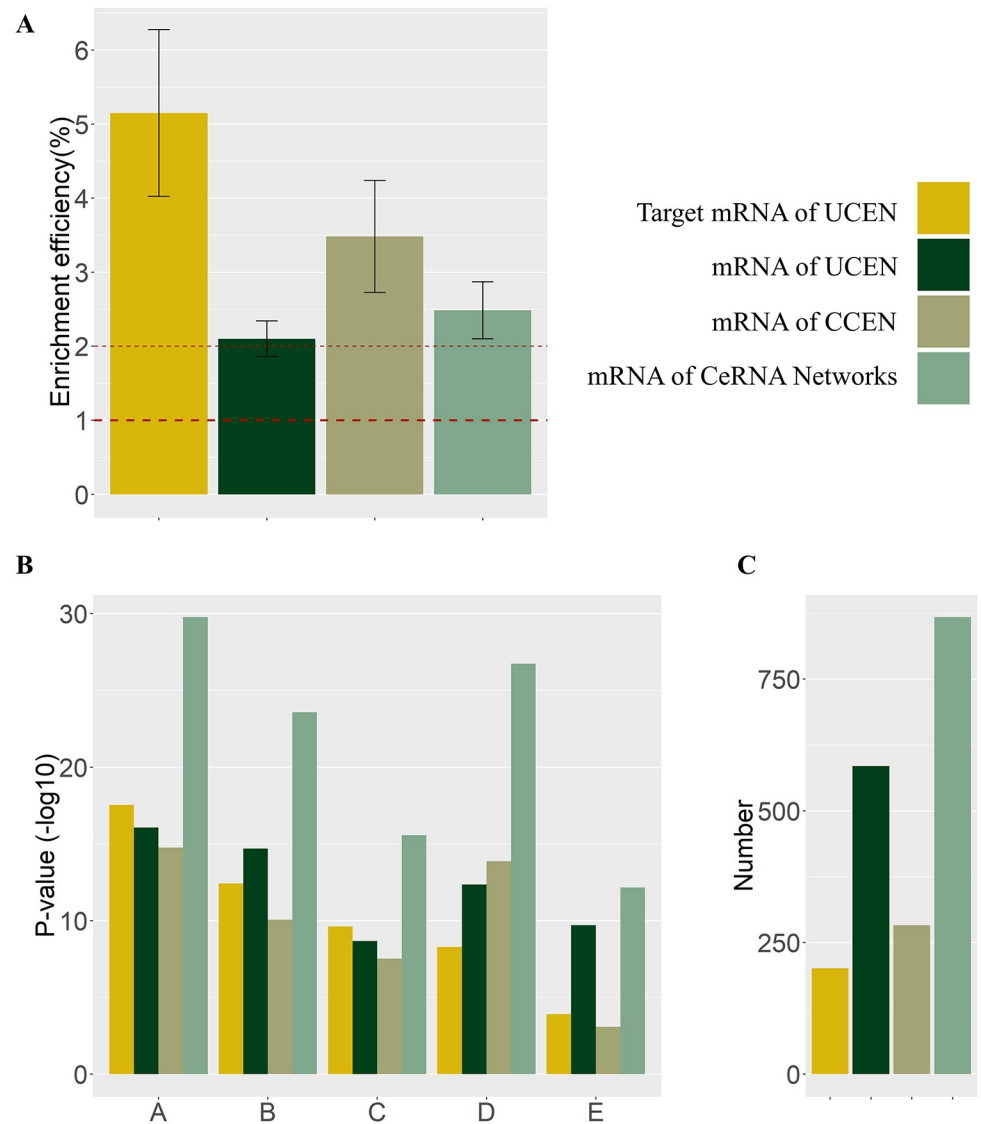


Fig 6. Results of enrichment efficiency analysis. (A) Values of enrichment efficiency of the mRNAs associated with microRNAs in the UCEN and those of direct analysis of mRNAs in CCEN, UCEN, and CeRNA networks. (B) P-values of enrichment analysis based on gene ontology and KEGG, where columns A, B, C, D and E correspond to blood vessel development, regulation of cell adhesion, circulatory system process, tissue morphogenesis, and pathways in cancer, respectively. (C) The numbers of mRNAs associated with microRNAs in UCEN and of all mRNAs in the CCEN, UCEN and CeRNA networks.

<https://doi.org/10.1371/journal.pcbi.1007793.g006>

mir-9, hsa-mir-21, hsa-mir-31, hsa-mir-148a, hsa-mir-195, and hsa-mir-375. For each stage, there is a specific microRNA that affects the survival rate: hsa-mir-19a (P-value, 0.048), hsa-mir-196b (P-value, 0.058), hsa-mir-194 (P-value, 0.061), and hsa-mir-144 (P-value, 0.012) for stages I-IV, respectively, as shown in Fig 8. To further support for this finding, we also calculate the four microRNAs survival curves of LUAD patients at each stage, and find that hsa-mir-19a, hsa-mir-196b, hsa-mir-194 and hsa-mir-144 have little effect on the survival at other stages except for stages I (S1 Fig), II (S2 Fig), III (S3 Fig) and IV (S4 Fig), respectively.

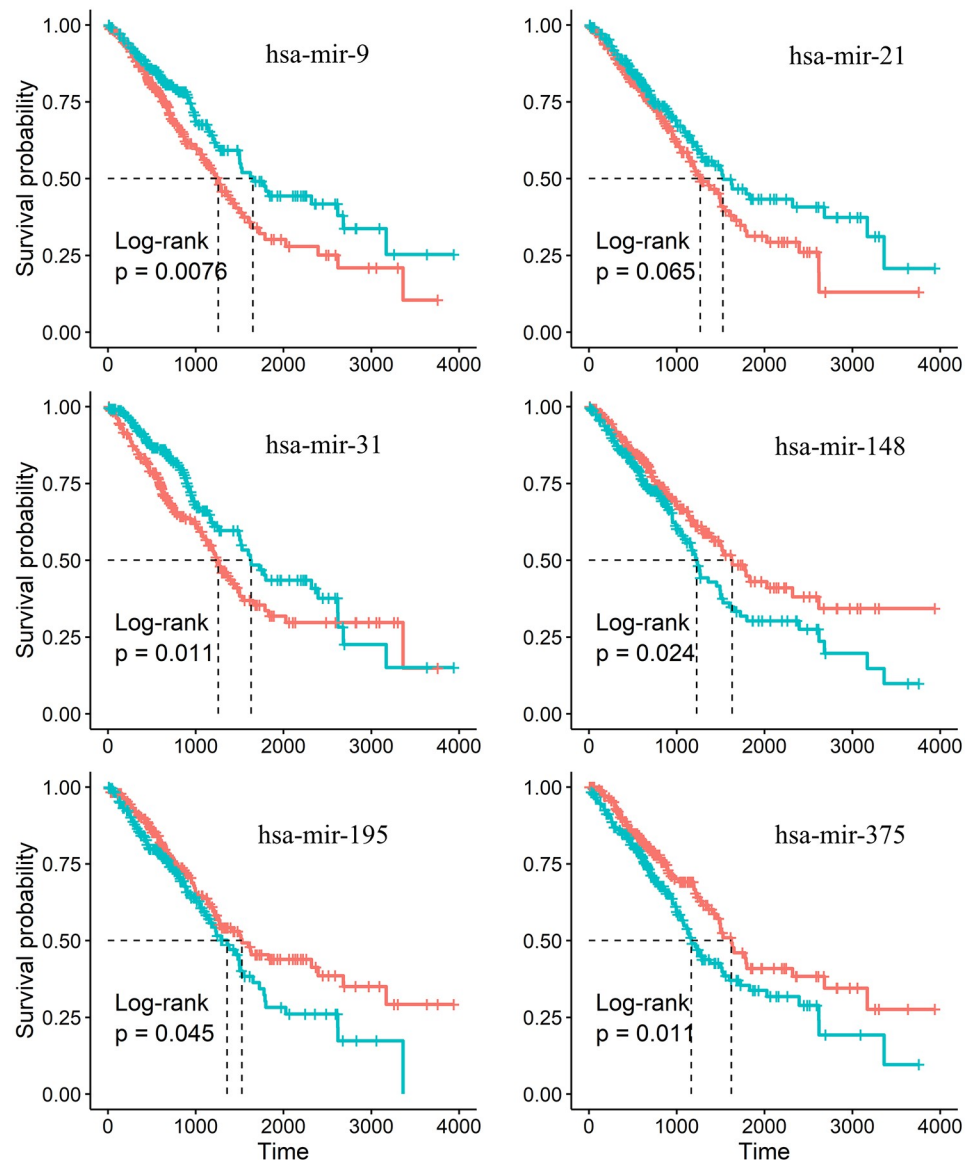


Fig 7. Survival curves of microRNAs from Kaplan-Meier analysis. Six microRNAs in CCEN that have a significant effect on the survival of LUAD patient survival in all four stages. Log-rank tests are used to analyze the Kaplan-Meier survival curve.

<https://doi.org/10.1371/journal.pcbi.1007793.g007>

Conclusion

LUAD has more pronounced genomic variations [40] than other lung cancer subtypes, which are rarely caused by a few genetic changes, rendering necessary articulating alternative methodologies in order to obtain a reasonable understanding of the complicated mechanisms behind the evolution of LUAD. Analyses based on the CeRNA hypothesis provide a promising approach to understanding mutual regulation at the post-transcriptional level [9, 10, 41].

We have reconstructed the CeRNA networks for lncRNAs, microRNAs and mRNAs corresponding to the four stages of LUAD. Support for the validity of the microRNA-associated CeRNA networks is obtained by comparing variations in the fold-change values and expression levels, as well as by enriching mRNAs to cancer-related items in gene ontology and

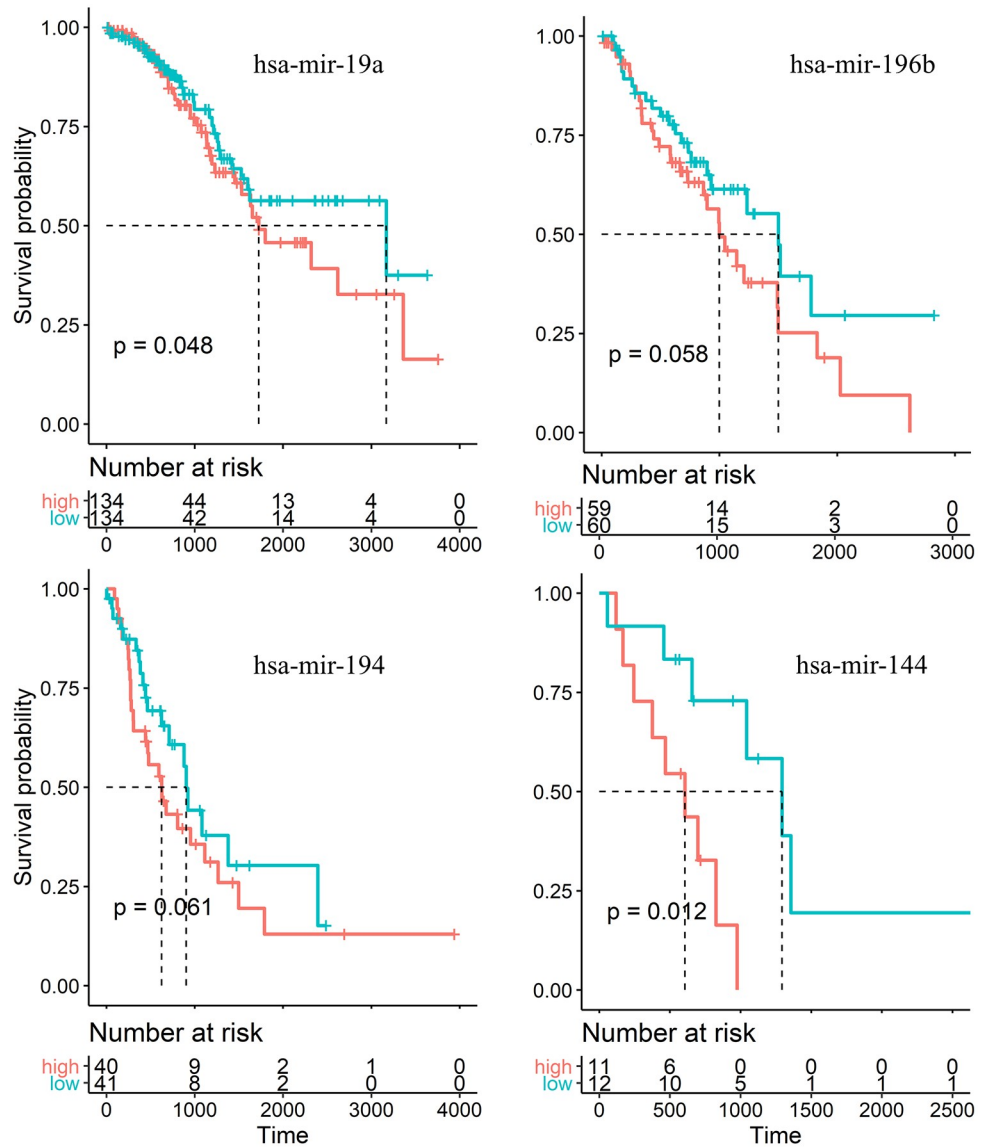


Fig 8. Survival curves of microRNAs from Kaplan-Meier analysis. Four microRNAs in the CeRNA networks, each affecting the survival of a specific stage. Log-rank tests are used to analyze the Kaplan-Meier survival curve.

<https://doi.org/10.1371/journal.pcbi.1007793.g008>

KEGG pathways. Through a systematic enrichment efficiency analysis on mRNAs linked by microRNA in the UCENs and mRNAs selected directly from CeRNA networks (Fig 6), we have identified the key underpinning RNA molecules for LUAD. Especially, by partitioning the network, we find that some specific microRNAs in the CCEN and UCENs have significant influence on the survival rate of LUAD patients (Figs 7 and 8). By performing KEGG pathway enrichment analysis on the mRNA of the CeRNA networks, we find that these mRNAs are not only related to cancer pathways as well as cell cycles and microRNAs in cancer, but there is also a reduction in the degree of the corresponding correlation as the patient state deteriorates from stage I to stage IV of LUAD. This is strong indication that these mRNAs and their CeRNA networks are involved in LUAD and its evolution.

Through the Kaplan-Meier survival analysis [39], we find six microRNAs that markedly affect the patient survival rate for all four stages of LUAD. The first is hsa-mir-9 reported in recent studies, which is involved in the regulation of NSCLC-related eukaryotic translation initiation factor 5A2, TGFBR2, Lamins and other biomolecules [42–44]. Another microRNA is hsa-mir-21, the first oncogenic microRNA discovered [45], which is associated with pulmonary fibrosis [46], inhibits PTEN, promotes growth and invasion [47], affects cell proliferation, migration, and survival [48, 49]. The third one is hsa-mir-31, which is involved in the inhibition of specific tumor suppressants in human lung cancer [50], affects the survival of LUAD patients [51] and participates in the regulation of lung cancer treatment [52, 53]. The fourth one is hsa-mir-148a, which inhibits the metastasis, proliferation and pathogenesis of NSCLC [54–56]. The fifth one is hsa-mir-195, which participates in the process of affecting lung cancer by regulating MYB, cyclin D3, Ailanthone [57–59]. The sixth one is hsa-mir-375, which affects YAP1 molecule [60] in lung cancer and is involved in the regulation of multiple lung cancer stages [61] as a target of Claudin-1 in NSCLC [62]. It is also involved in mouse alveoli inhibition of Wnt/b-catenin pathway [63]. These previous studies suggest that the six microRNAs that we have successfully identified through our quantitative analysis of the CeRNA networks of LUAD are indeed significant for cancer development, validating and demonstrating the power of our framework of dynamical network analysis.

We have also identified four microRNAs that affect the survival of patients at a specific stage of LUAD. In particular, hsa-mir-19a affects the survival at the first stage, which regulates biological molecules such as grape seed procyanidin, MTUS1, c-Met, and FOXP1 to affect lung cancer [64–67]. MicroRNA hsa-mir-196b influences the survival of patients at the second stage of LUAD, which regulates targets Homeobox A9 and Runx2, etc. [68, 69] and acts on early stage lung cancer [70]. MicroRNA hsa-mir-194 has an effect on the survival of patients at the third stage of LUAD, which can serve as a target for human nuclear distribution protein C [71] and constitute a negative feedback loop with Cullin 4B in carcinogenesis [72]. Finally, hsa-mir-144 impacts the patient survival at the fourth stage of LUAD, which regulates EZH2, TIGAR, Lico A, etc. that affect lung cancer [73–75].

The emerging field of network physiology and medicine [23–27] focuses on the associations between network structures and physiologic states, which relies heavily on measurements from biomedical experiments on large spatial and temporal scales. In this sense, our findings demonstrate an association between CeRNA network and physiologic dysfunction of the lung organ.

While CeRNA network analysis can play a pivotal role in the study of post-transcriptional gene regulation in cancer and disease treatment, there are limitations. For example, for the analysis to be valid, the relative concentrations of CeRNAs in the cellular microenvironment should not be too heterogeneous. There can also be defects and errors in the existing microRNA targets data bases.

Another issue concerns the use of possible surrogate test, which is useful for assessing the effects of different data sources on the results. However, our work is to analyze the CeRNA network constructed from genomics data collected from TCGA and clinical data from a large number of LUAD samples in a step-by-step manner. If we use surrogate test to obtain paired signals from different LUAD samples to construct the CeRNA network, there will be many possible CeRNA networks, making it extremely difficult (if not impossible) to determine the networks for identifying biomarkers. Another aspect of our work is a focus on the variations among the progressive stages of lung cancer, and the number of sequential CeRNA networks constructed for different LUAD stages is limited. Our approach is then to use the gene expression, the FC values of the CeRNA network nodes, and gene enrichment analysis to validate the network. For surrogate networks, the meanings of these tests are not clear. We have

demonstrated that our approach does lead to a number of microRNA biomarkers, and we have carried out indirect validation to test the effectiveness of the targets, which includes the Kaplan-Meier survival analysis based on combining the time event correspondence to the clinical physiological data of the sample patient and *a priori* knowledge or expert system verification.

Taken together, based on the CeRNA hypothesis, we have developed a framework to reconstruct the CeRNA networks for the four evolutionary stages of LUAD. Our analysis has yielded some prognostic markers closely related to the survival of LUAD patients. The comparative study of the CeRNA networks for the four stages of LUAD provides new insights into understanding cancer mechanisms and identifying targets for better drugs.

Materials and methods

Clinical and gene expression data of LUAD

A total of 877 samples with the corresponding lncRNA, microRNA and mRNA expression profile data from the cancer genome atlas (TCGA) were used. Especially, the numbers of lncRNA, microRNA and mRNA profiles are 29095, 1881 and 25527, respectively. Our analysis requires pre-processing to match the RNA expression data with the patient clinical data, which has excluded 269 samples. We follow the Tumor-Node-Metastasis (TNM) staging criteria for malignant tumors to classify LUAD into four stages. In particular, TNM is a standardized classification system established by the International Association for the Study of Lung Cancer to describe the development of lung cancer in terms of size and spread, where “T” describes the size of the tumor and any spread of the cancer into nearby tissues, “N” denotes the spread of the cancer to nearby lymph nodes, and “M” stands for metastasis, i.e., the spread of the cancer to other parts of the body [76]. For our datasets, we have that, of the remaining 608 samples, 288, 124, 85, and 26 are labeled as stages I, II, III, and IV, respectively, whereas 85 are normal. (More details about the classification criteria of LUAD can be found in [S1 Appendix](#)). We perform the match pre-processing on the gene expression profiles as well, resulting in 14460, 1881 and 24991 lncRNA, microRNA and mRNA profiles, respectively. We integrate and further match the clinical physiological data of the samples with either the lncRNA and mRNA data or the microRNA data. For match with the lncRNA and mRNA data, we have a total of 575 cases, where 26, 284, 122, 84, and 59 samples belong to stages I, II, III, IV, and normal, respectively. For match with the microRNA data, there are 24, 279, 122, 85, and 46 samples corresponding to stages I-IV LUAD and normal cases, respectively. (More details are listed in [S2 Table](#)).

Framework of analysis

Our articulated framework of dynamical network analysis combines the following methods of analysis: differential expression analysis, match with microRNA-mRNA and microRNA-lncRNA interaction data, identification of negative correlation between microRNA-mRNA and microRNA-lncRNA expressions, CeRNA network partition, gene ontology and KEGG enrichment analysis, and survival analysis. A flow chart of these methods is illustrated in [Fig 9](#). In the following, each of the methods is described.

Differential expression analysis

We use the R-language “Limma” package [77] to analyze the differentially expressed lncRNA, microRNA and mRNA profile data [78] in the four stages of LUAD, with the threshold of $\log_2 FC$ (\log_2 fold change) absolute value for filtering the three types RNAs set as one, while ensuring that their P-value (t-statistic [79], see [S2 Appendix](#) for detail) is less than 0.05. Fold changes

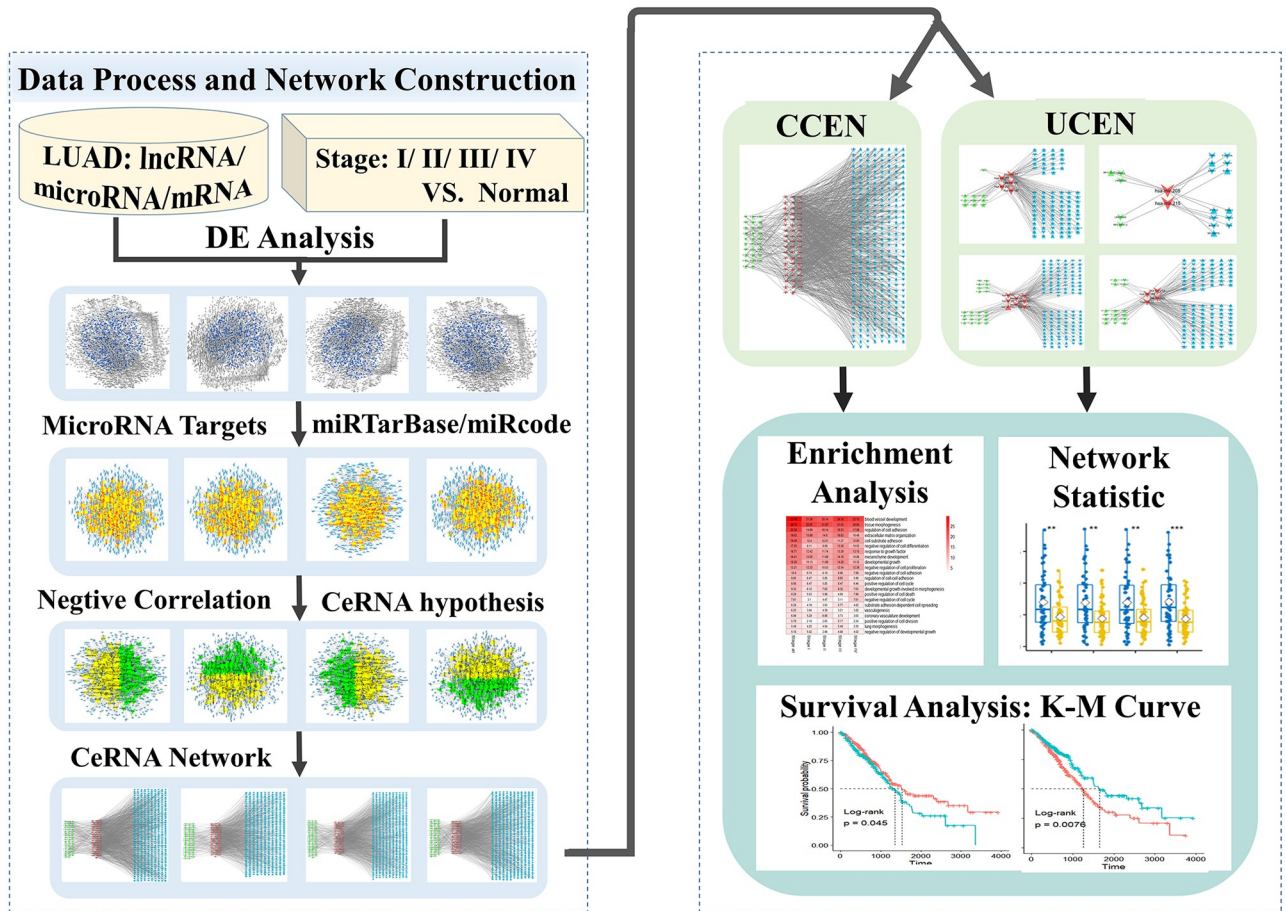


Fig 9. Overall work flow chart of proposed framework of dynamical network analysis of lung cancer. LUAD samples are divided into four stages according to the TNM criteria. Differential expression gene analysis and target selection based on negative correlation are performed on the gene expression data of lncRNAs, microRNAs and mRNAs to construct the corresponding CeRNA network for each of the four LUAD stages. The CeRNA networks are divided into UCENs and CCEN. A statistic analysis of the network nodes as well as gene enrichment and survival analyses are carried out.

<https://doi.org/10.1371/journal.pcbi.1007793.g009>

greater than one correspond to an up-regulated gene, while those less than minus one are associated with a down-regulated gene. We then obtain the numbers of lncRNAs, microRNAs and mRNAs and the up or down regulated expression genes in the four stages of LUAD, as listed in [S3 Table](#).

The individual RNA behavior can be assessed from the volcano map of RNA expression. Fold Change characterizes the relative expression level of interested samples to that of the control samples. For \log_2 Fold-Change = 1, the individual RNA expression level equals the group average of RNAs. We have provided the volcano maps for the expressions of lncRNA ([S5 Fig](#)), microRNA ([S6 Fig](#)) and mRNA ([S7 Fig](#)) at the four LUAD stages.

Data of microRNA-mRNA and microRNA-lncRNA interactions

The base of complementary pairing matching relationship data in the RNA sequences for microRNA-mRNA and microRNA-lncRNA are downloaded from the websites of TarBase (version 6.0) [80], miRTarBase (version 6.1) [81], and miRecords (version 4) [82]. The differentially expressed microRNAs are matched with the microRNA-mRNA and microRNA-lncRNA interaction relationships, following which the lncRNAs and mRNAs targeted by these

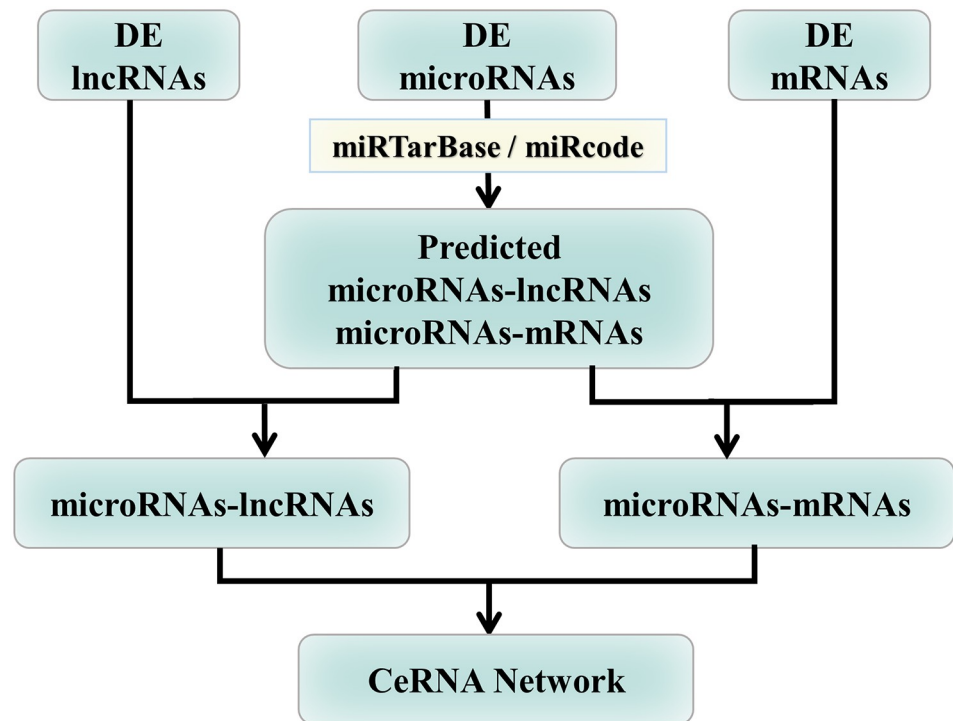


Fig 10. Selection of targets of microRNAs matched lncRNA and mRNA gene expression data. Differentially expressed microRNAs are matched to microRNAs of the miRTarBase and miRcode data bases to identify the lncRNAs and mRNAs of the microRNAs targets, which are then matched to differentially expressed lncRNAs and mRNAs, yielding the relationships between microRNAs-lncRNAs and microRNAs-mRNAs and, consequently, leading to the lncRNAs-microRNAs-mRNAs CeRNA networks for the four stages of LUAD.

<https://doi.org/10.1371/journal.pcbi.1007793.g010>

microRNAs are selected for matching the differentially expressed lncRNAs and mRNAs. A detailed workflow illustrating the analysis of microRNA-mRNA and microRNA-lncRNA interactions is shown in Fig 10, with more details in S4 Table.

Negative correlation between microRNA-mRNA and microRNA-lncRNA expressions

Based on the CeRNA hypothesis [17, 18], we pick out only those microRNA-mRNA and microRNA-lncRNA interaction pairs with negative correlation calculated from the Pearson correlation of the RNA expression data, where the threshold in the correlation coefficient is set to be -0.1. (See S5 Table for more details).

Partition of CeRNA network

We select the identical microRNAs among the LUAD four stages as well as the mRNAs and lncRNAs connected by these microRNAs in the CeRNA network as constituting the CCEN, where the numbers of identical lncRNAs, microRNAs and mRNAs emerging in the LUAD CeRNA networks of all four stages are 29, 53 and 283, respectively.

Correspondingly, we compare the one-of-a-kind microRNAs of a specified LUAD stage with those in other stages as well as the mRNAs and lncRNAs connected by those microRNAs in the CeRNA networks, which constitute the UCEN. The numbers of (lncRNAs, microRNAs, mRNAs) in the UCEN networks associated with stages I-IV are (27, 5, 197), (30, 2, 236), (41, 8, 312), and (36, 4, 335), respectively.

Gene ontology and KEGG enrichment analysis

We perform gene ontology and KEGG enrichment analysis for mRNAs of the LUAD CeRNA networks, taking into account the various biological processes associated with gene ontology.

Gene enrichment analysis is a widely used approach to identifying biological connections. In Figs 3 and 5, we implement the hypergeometric model to assess whether the number of selected genes in the CeRNA networks associated with lung cancer is larger than that which can be expected purely by chance. In particular, the P-value determines whether any term annotates a specified list of genes at frequency greater than that which can be expected by chance, as determined by the hypergeometric distribution:

$$p = 1 - \sum_{i=0}^{k-1} \frac{\binom{N-M}{n-i} \binom{M}{i}}{\binom{N}{n}}, \quad (1)$$

where N is the total number of genes in the background distribution, M is the number of genes within that distribution that are annotated (either directly or indirectly) to the node of interest, n is the size of the list of genes of interest, and k is the number of genes within that list which are annotated to the node. The background distribution by default is all the genes that have an annotation. Then, the outcome of a statistical hypothesis test for the hypergeometric distribution in the gene enrichment analysis, divided by the total number N of genes in the analysis, defines the enrichment efficiency η :

$$\eta = -\log_{10} P/N. \quad (2)$$

Survival analysis

We use the Kaplan-Meier method [39], a non-parametric method, to estimate the survival probability from the observed survival data. The survival probability as a function of time is calculated according to

$$S(t_i) = S(t_{i-1})(1 - d_i/n_i), \quad (3)$$

where n_i is the number of patients who were alive before time t_i and d_i is the number of death events at t_i . The estimated probability is a step function that changes value only at the time of each event.

Log-rank tests are used to carry out a univariate analysis of the Kaplan-Meier survival curve, which belong to the chi-square test, where all time points of the sample survival information are equal (i.e., with weight setting to one).

Computational packages

We carry out data processing and statistical analysis using R language (v.3.5.1), analyze differentially expressed RNAs by using the “Limma” package [77], and plot the heatmap of the enrichment analysis and the Kaplan-Meier survival curves using the “heatmap,” “survival,” and “survminer” packages. The various CeRNA networks are visualized via Cytoscape (v3.6.1). Finally, we perform the gene ontology and KEGG functional enrichment analysis using the online tool Metascape (<http://metascape.org>).

Supporting information

S1 Appendix. TNM stages.
(PDF)

S2 Appendix. Statistical analysis and P-value. We have used four methods of statistical analysis involving hypothesis testing to calculate the P-values.

(PDF)

S1 Fig. Survival curves of hsa-mir-19a. The P-values of hsa-mir-19a in the Kaplan-Meier survival curves are 0.24, 0.95, 0.13 for stages II, III and IV, respectively. Log-rank tests are used to analysis of Kaplan-Meier survival curve.

(TIF)

S2 Fig. Survival curves of hsa-mir-196b. The P-values of hsa-mir-196b in the Kaplan-Meier survival curves are 0.58, 0.7, 0.99 for stages I, III and IV, respectively.

(TIF)

S3 Fig. Survival curves of hsa-mir-194. The P-values of hsa-mir-194 in the Kaplan-Meier survival curves are 0.87, 0.92, 0.6 for stages I, II and IV, respectively.

(TIF)

S4 Fig. Survival curves of hsa-mir-144. The P-values of hsa-mir-144 in the Kaplan-Meier survival curves are 0.9, 0.35, 0.58 for stages I, II and III, respectively.

(TIF)

S5 Fig. Individual lncRNA expressions. Volcano map of the lncRNA expression level of four stages of LUAD samples. The x -axis is \log_2 Fold-Change (FC value), and the y -axis is the P-values from the differential expression analysis.

(TIF)

S6 Fig. Individual microRNA expressions. Volcano map of the microRNA expression level of four stages of LUAD samples.

(TIF)

S7 Fig. Individual mRNA expressions. Volcano map of the mRNA expression level of four stages of LUAD samples.

(TIF)

S1 Table. The number of nodes and edges of CeRNA networks. Detailed parameter values of the reconstructed lncRNA-microRNA-mRNA CeRNA networks.

(PDF)

S2 Table. Distribution of clinical samples in gene expression data of LUAD. Clinical physiological data of the four LUAD stages matched with the lncRNA and mRNA expression data.

(PDF)

S3 Table. The number of differential expression (DE) gene. Differentially expressed lncRNA, microRNA and mRNA profile data in the four stages of LUAD.

(PDF)

S4 Table. The number of lncRNAs or mRNAs targeted by microRNAs and their relationships. The microRNA-mRNA and microRNA-lncRNA interactions relationships obtained by using the base of complementary pairing matching relationship data in the RNA sequences.

(PDF)

S5 Table. The number of lncRNAs or mRNAs selected by introducing negative correlation. The Pearson correlation of the RNA expression data between microRNA-mRNA and microRNA-lncRNA expressions is calculated.

(PDF)

S1 Data. Figure data.
(RAR)

Acknowledgments

We thank Prof. Celso Grebogi and Prof. Lei Yang for helpful discussions.

Author Contributions

Conceptualization: Zhi-Tong Bing, Bing-Hui Guo, Liang Huang, Zi-Gang Huang, Ying-Cheng Lai.

Data curation: Chao Kong, Zhi-Tong Bing.

Formal analysis: Chao Kong, Zhi-Tong Bing, Bing-Hui Guo, Zi-Gang Huang.

Funding acquisition: Zhi-Tong Bing, Bing-Hui Guo, Zi-Gang Huang.

Investigation: Chao Kong, Yu-Xiang Yao, Zhi-Tong Bing.

Methodology: Yu-Xiang Yao, Zhi-Tong Bing, Zi-Gang Huang.

Project administration: Zhi-Tong Bing, Bing-Hui Guo, Zi-Gang Huang.

Resources: Zhi-Tong Bing, Bing-Hui Guo, Liang Huang, Zi-Gang Huang.

Validation: Chao Kong, Zi-Gang Huang, Ying-Cheng Lai.

Visualization: Chao Kong.

Writing – original draft: Chao Kong.

Writing – review & editing: Yu-Xiang Yao, Liang Huang, Zi-Gang Huang, Ying-Cheng Lai.

References

1. Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-Tieulent J, Jemal A. Global cancer statistics, 2012. *CA Cancer J Clin.* 2015; 65(2):87–108. <https://doi.org/10.3322/caac.21262> PMID: 25651787
2. Segal R, Miller K, Jemal A. Cancer statistics, 2018. *CA Cancer J Clin.* 2018; 68:7–30. <https://doi.org/10.3322/caac.21442>
3. Molina JR, Yang P, Cassivi SD, Schild SE, Adjei AA. Non-Small Cell Lung Cancer: Epidemiology, Risk Factors, Treatment, and Survivorship. *Mayo Clin Proc.* 2008; 83(5):584–594. <https://doi.org/10.4065/83.5.584> PMID: 18452692
4. Seo JS, Ju YS, Lee WC, Shin JY, Lee JK, Bleazard T, et al. The transcriptional landscape and mutational profile of lung adenocarcinoma. *Geno Res.* 2012; 22(11):2109–2119. <https://doi.org/10.1101/gr.145144.112>
5. Network CGAR, Others. Comprehensive molecular profiling of lung adenocarcinoma. *Nature.* 2014; 511(7511):543. <https://doi.org/10.1038/nature13385>
6. Epsi NJ, Panja S, Pine SR, Mitrofanova A. pathCHEMO, a generalizable computational framework uncovers molecular pathways of chemoresistance in lung adenocarcinoma. *Commun Biol.* 2019; 2(1):1–13. <https://doi.org/10.1038/s42003-019-0572-6>
7. Baek D, Villén J, Shin C, Camargo FD, Gygi SP, Bartel DP. The impact of microRNAs on protein output. *Nature.* 2008; 455(7209):64. <https://doi.org/10.1038/nature07242> PMID: 18668037
8. Bartel DP. MicroRNAs: target recognition and regulatory functions. *Cell.* 2009; 136(2):215–233. <https://doi.org/10.1016/j.cell.2009.01.002> PMID: 19167326
9. Tay Y, Karreth FA, Pandolfi PP. Aberrant ceRNA activity drives lung cancer. *Cell Res.* 2014; 24(3):259. <https://doi.org/10.1038/cr.2014.21> PMID: 24525785
10. Adams BD, Kasinski AL, Slack FJ. Aberrant Regulation and Function of MicroRNAs in Cancer. *Current Biology.* 2014; 24(16):R762–R776. <https://doi.org/10.1016/j.cub.2014.06.043> PMID: 25137592
11. Consortium EP, Others. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature.* 2007; 447(7146):799. <https://doi.org/10.1038/nature05874>

12. Jiang Q, Wang Y, Hao Y, Juan L, Teng M, Zhang X, et al. miR2Disease: A manually curated database for microRNA deregulation in human disease. *Nucleic Acids Research*. 2009; 37:98–104. <https://doi.org/10.1093/nar/gkn714>
13. Mercer TR, Dinger ME, Mattick JS. Long non-coding RNAs: insights into functions. *Nat Rev Gene*. 2009; 10(3):155. <https://doi.org/10.1038/nrg2521>
14. Xu J, Li CX, Lv JY, Li YS, Xiao Y, Shao TT, et al. Prioritizing candidate disease miRNAs by topological features in the miRNA target–dysregulated network: Case study of prostate cancer. *Mole Cancer Ther-apeu*. 2011; 10(10):1857–1866. <https://doi.org/10.1158/1535-7163.MCT-11-0055>
15. Li Y, Qiu C, Tu J, Geng B, Yang J, Jiang T, et al. HMDD v2.0: a database for experimentally supported human microRNA and disease associations. *Nucleic Acids Research*. 2014; 42:1070–1074. <https://doi.org/10.1093/nar/gkt1023>
16. Anastasiadou E, Jacob LS, Slack FJ. Non-coding RNA networks in cancer. *Nat Rev Cancer*. 2018; 18(1):5. <https://doi.org/10.1038/nrc.2017.99> PMID: 29170536
17. Ebert MS, Neilson JR, Sharp PA. MicroRNA sponges: competitive inhibitors of small RNAs in mammalian cells. *Nat Meth*. 2007; 4(9):721. <https://doi.org/10.1038/nmeth1079>
18. Salmena L, Poliseno L, Tay Y, Kats L, Pandolfi PP. A ceRNA hypothesis: the Rosetta Stone of a hidden RNA language? *Cell*. 2011; 146(3):353–358. <https://doi.org/10.1016/j.cell.2011.07.014> PMID: 21802130
19. Tay Y, Rinn J, Pandolfi PP. The multilayered complexity of ceRNA crosstalk and competition. *Nature*. 2014; 505(7483):344. <https://doi.org/10.1038/nature12986> PMID: 24429633
20. Shao T, Wu A, Chen J, Chen H, Lu J, Bai J, et al. Identification of module biomarkers from the dysregulated ceRNA–ceRNA interaction network in lung adenocarcinoma. *Mole Biosys*. 2015; 11(11):3048–3058. <https://doi.org/10.1039/C5MB00364D>
21. Robinson DR, Wu YM, Lonigro RJ, Vats P, Cobain E, Everett J, et al. Integrative clinical genomics of metastatic cancer. *Nature*. 2017; 548(7667):297. <https://doi.org/10.1038/nature23306> PMID: 28783718
22. Network CGAR, et al. Comprehensive genomic characterization of squamous cell lung cancers. *Nature*. 2012; 489(7417):519. <https://doi.org/10.1038/nature11404>
23. Ivanov PC, Liu K, Bartsch RP. Focus on the emerging new fields of network physiology and network medicine. *New Journal of Physics*. 2016; 18(10):100201. <https://doi.org/10.1088/1367-2630/18/10/100201> PMID: 30881198
24. Bartsch RP, Liu KK, Bashan A, Ivanov PC. Network physiology: how organ systems dynamically interact. *PLOS ONE*. 2015; 10(11):e0142143. <https://doi.org/10.1371/journal.pone.0142143> PMID: 26555073
25. Bashan A, Bartsch RP, Kantelhardt JW, Havlin S, Ivanov PC. Network physiology reveals relations between network topology and physiological function. *Nat Commun*. 2012; 3(1):1–9.
26. Liu KK, Bartsch RP, Lin A, Mantegna RN, Ivanov PC. Plasticity of brain wave network interactions and evolution across physiologic states. *Front Neural Circ*. 2015; 9:62.
27. Liu KK, Bartsch RP, Ma QD, Ivanov PC; IOP Publishing. Major component analysis of dynamic networks of physiologic organ interactions. *J Phys Conf Ser*. 2015; 640(1):012013. <https://doi.org/10.1088/1742-6596/640/1/012013> PMID: 30174717
28. Bascompte J, Jordano P, Melián CJ, Olesen JM. The nested assembly of plant-animal mutualistic networks. *Proc Natl Acad Sci (USA)*. 2003; 100:9383–9387. <https://doi.org/10.1073/pnas.1633576100>
29. Guimaraes PR, Jordano P, Thompson JN. Evolution and coevolution in mutualistic networks. *Ecol Lett*. 2011; 14:877–885. <https://doi.org/10.1111/j.1461-0248.2011.01649.x> PMID: 21749596
30. Lever JJ, Nes EH, Scheffer M, Bascompte J. The sudden collapse of pollinator communities. *Ecol Lett*. 2014; 17(3):350–359. <https://doi.org/10.1111/ele.12236> PMID: 24386999
31. Rohr RP, Saavedra S, Bascompte J. On the structural stability of mutualistic systems. *Science*. 2014; 345(6195):1253497. <https://doi.org/10.1126/science.1253497> PMID: 25061214
32. Dakos V, Bascompte J. Critical slowing down as early warning for the onset of collapse in mutualistic communities. *Proc Natl Acad Sci (USA)*. 2014; 111(49):17546–17551. <https://doi.org/10.1073/pnas.1406326111>
33. Guimaraes PR, Pires MM, Jordano P, Bascompte J, Thompson JN. Indirect effects drive coevolution in mutualistic networks. *Nature*. 2017; 550:511–514. <https://doi.org/10.1038/nature24273> PMID: 29045396
34. Jiang J, Huang ZG, Seager TP, Lin W, Grebogi C, Hastings A, et al. Predicting tipping points in mutualistic networks through dimension reduction. *Proc Natl Acad Sci (USA)*. 2018; 115:E639–E647. <https://doi.org/10.1073/pnas.1714958115>

35. Jiang J, Hastings A, Lai YC. Harnessing tipping points in complex ecological networks. *Roy Soc J Interface*. 2019; 16:20190345. <https://doi.org/10.1098/rsif.2019.0345>
36. Ohgushi T, Schmitz O, Holt RD. *Trait-Mediated Indirect Interactions: Ecological and Evolutionary Perspectives*. Cambridge UK: Cambridge Univ. Press; 2012.
37. Bauer DF. Constructing confidence sets using rank statistics. *J Ame Stat Asso*. 1972; 67:687–690. <https://doi.org/10.1080/01621459.1972.10481279>
38. Hollander M, Wolfe DA. *Nonparametric Statistical Methods*. New York: John Wiley & Sons; 1973.
39. Kaplan EL, Meier P. Nonparametric estimation from incomplete observations. *J Ame Stat Asso*. 1958; 53(282):457–481. <https://doi.org/10.1080/01621459.1958.10501452>
40. Devarakonda S, Morgensztern D, Govindan R. Genomic alterations in lung adenocarcinoma. *Lancet Oncol*. 2015; 16(7):e342–e351. [https://doi.org/10.1016/S1470-2045\(15\)00077-7](https://doi.org/10.1016/S1470-2045(15)00077-7) PMID: 26149886
41. Li H, Chen S, Liu J, Guo X, Xiang X, Dong T, et al. Long non-coding RNA PVT1-5 promotes cell proliferation by regulating miR-126/SLC7A5 axis in lung cancer. *Biochem Biophys Res Commun*. 2018; 495(3):2350–2355. <https://doi.org/10.1016/j.bbrc.2017.12.114> PMID: 29277611
42. Pan Q, Sun L, Zheng D, Li N, Shi H, Song J, et al. MicroRNA-9 Enhanced Cisplatin Sensitivity in Non-small Cell Lung Cancer Cells by Regulating Eukaryotic Translation Initiation Factor 5A2. *BioMed Res Int*. 2018; 2018:1769040. <https://doi.org/10.1155/2018/1769040> PMID: 30175116
43. Li G, Wu F, Yang H, Deng X, Yuan Y. MiR-9-5p promotes cell growth and metastasis in non-small cell lung cancer through the repression of TGFBR2. *Biomed Pharmacothera*. 2017; 96:1170–1178. <https://doi.org/10.1016/j.biopha.2017.11.105>
44. Guinde J, Frankel D, Perrin S, Delecourt V, Lévy N, Barlesi F, et al. Lamins in Lung Cancer: Biomarkers and Key Factors for Disease Progression through miR-9 Regulation? *Cell*. 2018; 7(7):78. <https://doi.org/10.3390/cells7070078>
45. Kasinski AL, Slack FJ. MicroRNAs en route to the clinic: progress in validating and targeting microRNAs for cancer therapy. *Nat Rev Cancer*. 2011; 11(12):849. <https://doi.org/10.1038/nrc3166> PMID: 22113163
46. Liu G, Friggeri A, Yang Y, Milosevic J, Ding Q, Thannickal VJ, et al. miR-21 mediates fibrogenic activation of pulmonary fibroblasts and lung fibrosis. *J Exp Med*. 2010; 207(8):1589–1597. <https://doi.org/10.1084/jem.20100035> PMID: 20643828
47. Bica-Pop C, Cojocneanu-Petric R, Magdo L, Raduly L, Gulei D, Berindan-Neagoe I. Overview upon miR-21 in lung cancer: focus on NSCLC. *Cellu Mole Life Sci*. 2018; 75(19):3539–3551. <https://doi.org/10.1007/s00018-018-2877-x>
48. Zhu S, Wu H, Wu F, Nie D, Sheng S, Mo YY. MicroRNA-21 targets tumor suppressor genes in invasion and metastasis. *Cell Res*. 2008; 18(3):350. <https://doi.org/10.1038/cr.2008.24> PMID: 18270520
49. Markou A, Tsaroucha EG, Kaklamanis L, Fotinou M, Georgoulas V, Lianidou ES. Prognostic value of mature microRNA-21 and microRNA-205 overexpression in non-small cell lung cancer by quantitative real-time RT-PCR. *Clin Chem*. 2008; 54(10):1696–1704. <https://doi.org/10.1373/clinchem.2007.101741> PMID: 18719201
50. Liu X, Sempere LF, Ouyang H, Memoli VA, Andrew AS, Luo Y, et al. MicroRNA-31 functions as an oncogenic microRNA in mouse and human lung cancer cells by repressing specific tumor suppressors. *J Clin Invest*. 2010; 120(4):1298–1309. <https://doi.org/10.1172/JCI39566>
51. Meng W, Ye Z, Cui R, Perry J, Dedousi-Huebner V, Huebner A, et al. MicroRNA-31 predicts the presence of lymph node metastases and survival in patients with lung adenocarcinoma. *Clin Cancer Res*. 2013; 19(19):5423–5433. <https://doi.org/10.1158/1078-0432.CCR-13-0320> PMID: 23946296
52. Naidu S, Garofalo M. microRNAs: an emerging paradigm in lung cancer chemoresistance. *Front Med*. 2015; 2:77. <https://doi.org/10.3389/fmed.2015.00077>
53. Dong Z, Zhong Z, Yang L, Wang S, Gong Z. MicroRNA-31 inhibits cisplatin-induced apoptosis in non-small cell lung cancer cells by regulating the drug transporter ABCB9. *Cancer Lett*. 2014; 343(2):249–257. <https://doi.org/10.1016/j.canlet.2013.09.034> PMID: 24099915
54. Xie Q, Yu Z, Lu Y, Fan J, Ni Y, Ma L. microRNA-148a-3p inhibited the proliferation and epithelial–mesenchymal transition progression of non-small-cell lung cancer via modulating Ras/MAPK/Erk signaling. *J Cell Physiol*. 2018; 234(8):12786–12799. <https://doi.org/10.1002/jcp.27899> PMID: 30536836
55. Li J, Yu T, Cao J, Liu L, Liu Y, Kong HW, et al. MicroRNA-148a suppresses invasion and metastasis of human non-small-cell lung cancer. *Cell Physiol Biochem*. 2015; 37(5):1847–1856. <https://doi.org/10.1159/000438546> PMID: 26584284
56. Joshi P, Jeon YJ, Laganà A, Middleton J, Secchiero P, Garofalo M, et al. MicroRNA-148a reduces tumorigenesis and increases TRAIL-induced apoptosis in NSCLC. *Proc Nat Acad Sci (USA)*. 2015; 112(28):8650–8655. <https://doi.org/10.1073/pnas.1500886112>

57. Zhou Y, Tian L, Wang X, Ye L, Zhao G, Yu M, et al. MicroRNA-195 inhibits non-small cell lung cancer cell proliferation, migration and invasion by targeting MYB. *Cancer Lett.* 2014; 347(1):65–74. <https://doi.org/10.1016/j.canlet.2014.01.019>
58. Yu X, Zhang Y, Cavazos D, Ma X, Zhao Z, Du L, et al. miR-195 targets cyclin D3 and survivin to modulate the tumorigenesis of non-small cell lung cancer. *Cell Death Dis.* 2018; 9(2):193. <https://doi.org/10.1038/s41419-017-0219-9>
59. Hou S, Cheng Z, Wang W, Wang X, Wu Y. Ailanthone exerts an antitumor function on the development of human lung cancer by upregulating microRNA-195. *Journal of Cellular Biochemistry.* 2019; 120(6):10444–10451. <https://doi.org/10.1002/jcb.28329> PMID: 30565729
60. Nishikawa E, Osada H, Okazaki Y, Arima C, Tomida S, Tatematsu Y, et al. miR-375 is activated by ASH1 and inhibits YAP1 in a lineage-dependent manner in lung cancer. *Cancer Res.* 2011; 71(19):6165–6173. <https://doi.org/10.1158/0008-5472.CAN-11-1020> PMID: 21856745
61. Jin Y, Liu Y, Zhang J, Huang W, Jiang H, Hou Y, et al. The Expression of miR-375 Is Associated with Carcinogenesis in Three Subtypes of Lung Cancer. *PLOS ONE.* 2015; 10(12):e0144187. <https://doi.org/10.1371/journal.pone.0144187> PMID: 26642205
62. Yoda S, Soejima K, Hamamoto J, Yasuda H, Nakayama S, Satomi R, et al. Claudin-1 is a novel target of miR-375 in non-small-cell lung cancer. *Lung Cancer.* 2014; 85(3):366–372. <https://doi.org/10.1016/j.lungcan.2014.06.009> PMID: 25001509
63. Wang Y, Huang C, Reddy Chintagari N, Bhaskaran M, Weng T, Guo Y, et al. miR-375 regulates rat alveolar epithelial cell trans-differentiation by inhibiting Wnt/ β -catenin pathway. *Nuc Acids Res.* 2013; 41(6):3833–3844. <https://doi.org/10.1093/nar/gks1460>
64. Mao JT, Xu B, Smoake J, Lu QY, Park H, Henning SM, et al. MicroRNA-19a/b mediates grape seed procyanidin extract-induced anti-neoplastic effects against lung cancer. *J Nutri Biochem.* 2016; 34:118–125. <https://doi.org/10.1016/j.jnutbio.2016.05.003>
65. Gu Y, Liu S, Zhang X, Chen G, Liang H, Yu M, et al. Oncogenic miR-19a and miR-19b co-regulate tumor suppressor MTUS1 to promote cell proliferation and migration in lung cancer. *Prot Cell.* 2017; 8(6):455–466. <https://doi.org/10.1007/s13238-017-0393-7>
66. Cao X, Lai S, Hu F, Li G, Wang G, Luo X, et al. miR-19a contributes to gefitinib resistance and epithelial mesenchymal transition in non-small cell lung cancer cells by targeting c-Met. *Scientific Reports.* 2017; 7(1):2939. <https://doi.org/10.1038/s41598-017-01153-0> PMID: 28592790
67. Yamamoto K, Ito S, Hanafusa H, Shimizu K, Ouchida M. Uncovering Direct Targets of MiR-19a Involved in Lung Cancer Progression. *PLOS ONE.* 2015; 10(9):e0137887. <https://doi.org/10.1371/journal.pone.0137887> PMID: 26367773
68. Yu SL, Lee DC, Sohn HA, Lee SY, Jeon HS, Lee JH, et al. Homeobox A9 directly targeted by miR-196b regulates aggressiveness through nuclear Factor-kappa B activity in non-small cell lung cancer cells. *Mole Carcinog.* 2016; 55(12):1915–1926. <https://doi.org/10.1002/mc.22439>
69. Bai X, Meng L, Sun H, Li Z, Zhang X, Hua S. MicroRNA-196b inhibits cell growth and metastasis of lung cancer cells by targeting Runx2. *Cell Physiol Biochem.* 2017; 43(2):757–767. <https://doi.org/10.1159/000481559> PMID: 28950255
70. Tellez CS, Juri DE, Do K, Picchi MA, Wang T, Liu G, et al. miR-196b is epigenetically silenced during the premalignant stage of lung carcinogenesis. *Cancer Res.* 2016; 76(16):4741–4751. <https://doi.org/10.1158/0008-5472.CAN-15-3367> PMID: 27302168
71. Zhou L, Di Q, Sun B, Wang X, Li M, Shi J. MicroRNA-194 restrains the cell progression of non-small cell lung cancer by targeting human nuclear distribution protein C. *Oncol Rep.* 2016; 35(6):3435–3444. <https://doi.org/10.3892/or.2016.4708> PMID: 27035759
72. Mi J, Zou Y, Lin X, Lu J, Zhao H, Ye X, et al. Dysregulation of the miR-194–CUL4B negative feedback loop drives tumorigenesis in non-small-cell lung carcinoma. *Mole Oncol.* 2017; 11(3):305–319. <https://doi.org/10.1002/1878-0261.12038>
73. Liu C, Yang Z, Deng Z, Zhou Y, Gong Q, Zhao R, et al. Downregulated miR-144-3p contributes to progression of lung adenocarcinoma through elevating the expression of EZH2. *Cancer Med.* 2018; 7(11):5554–5566. <https://doi.org/10.1002/cam4.1714> PMID: 30280514
74. Chen S, Li P, Li J, Wang Y, Du Y, Chen X, et al. MiR-144 inhibits proliferation and induces apoptosis and autophagy in lung cancer cells by targeting TIGAR. *Cell Physiol Biochem.* 2015; 35(3):997–1007. <https://doi.org/10.1159/000369755> PMID: 25660220
75. Chen G, Ma Y, Jiang Z, Feng Y, Han Y, Tang Y, et al. Lico A Causes ER Stress and Apoptosis via Up-Regulating miR-144-3p in Human Lung Cancer Cell Line H292. *Frontiers in Pharmacology.* 2018; 9:837. <https://doi.org/10.3389/fphar.2018.00837> PMID: 30108506
76. Goldstraw P, Crowley J, Chansky K, Giroux DJ, Groome PA, Rami-Porta R, et al. The IASLC Lung Cancer Staging Project: proposals for the revision of the TNM stage groupings in the forthcoming (seventh)

- edition of the TNM Classification of malignant tumours. *J Thoracic Oncol.* 2007; 2(8):706–714. <https://doi.org/10.1097/JTO.0b013e31812f3c1a>
77. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research.* 2015; 43(7):e47. <https://doi.org/10.1093/nar/gkv007> PMID: 25605792
 78. Smyth G. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat Appl Genet Mol Biol.* 2004; 3:3. <https://doi.org/10.2202/1544-6115.1027>
 79. Phipson B, Lee S, Majewski IJ, Alexander WS, Smyth GK. Robust hyperparameter estimation protects against hypervariable genes and improves power to detect differential expression. *Ann Appl Stat.* 2016; 10(2):946. <https://doi.org/10.1214/16-AOAS920> PMID: 28367255
 80. Vergoulis T, Vlachos IS, Alexiou P, Georgakilas G, Maragkakis M, Reczko M, et al. TarBase 6.0: capturing the exponential growth of miRNA targets with experimental support. *Nucleic Acids Research.* 2012; 40:222–229. <https://doi.org/10.1093/nar/gkr1161>
 81. Hsu S, Tseng Y, Shrestha S, Lin Y, Khaleel A, Chou C, et al. miRTarBase update 2014: an information resource for experimentally validated miRNA-target interactions. *Nucleic Acids Research.* 2014; 42:78–85. <https://doi.org/10.1093/nar/gkt1266>
 82. Xiao F, Zuo Z, Cai G, Kang S, Gao X, Li T. miRecords: an integrated resource for microRNA-target interactions. *Nucleic Acids Research.* 2009; 37:105–110. <https://doi.org/10.1093/nar/gkn851>

Supplementary Information for
**Dynamical network analysis reveals key microRNAs in
progressive stages of lung cancer**

Chao Kong, Yu-Xiang Yao, Zhi-Tong Bing, Bing-Hui Guo, Liang Huang,
Zi-Gang Huang, and Ying-Cheng Lai

Corresponding author: Zi-Gang Huang

Contents

Supplementary Notes	1
Note 1: Survival curves	1
Note 2: TNM stages	1
Note 3: Statistical analysis and P-value	2
Note 4: Individual RNA expressions in the volcano map	2
Supplementary Figures	3
Supplementary Tables	7

Supplementary Notes

Note 1: Survival curves

Corroborating the results in Fig. 8 in the main text, we calculate the microRNA-based survival curves of LUAD patients at each stage. As described in the main text, microRNAs hsa-mir-19a, hsa-mir-196b, hsa-mir-194 and hsa-mir-144 affect the survival of LUAD patients in stages I, II, III and IV, respectively.

Further support for this finding is presented in Supplementary Figs. 1-4, which show that the four microRNAs have little effect on the survival of LUAD patients at other stages except for stages I, II, III and IV, respectively. More specifically, the P-values of hsa-mir-19a in the Kaplan-Meier survival curves are 0.24, 0.95, 0.13 for stages II, III and IV, respectively. For hsa-mir-196b, the P-values are 0.58, 0.7, 0.99 for stages I, III and IV, respectively. For hsa-mir-194, the P-values are 0.87, 0.92, 0.6 for stages I, II and IV, respectively. For hsa-mir-144, the P-values are 0.9, 0.35, 0.58 for stages I, II and III, respectively.

Note 2: TNM stages

Lung Adenocarcinoma (LUAD) is a kind of lung cancer that is a malignant tumor. The International Association for the Study of Lung Cancer established the Lung Cancer Staging Project in 1998 and updated the Tumor-Node-Metastasis (TNM) classification system of malignant tumors, which forms the basis for classifying lung cancer into distinct stages. The current (seventh) edition classifies all histotypes of lung cancer after evaluation of the outcomes from an extensive worldwide database of patients.

TNM characterizes the size and location of the tumor, the location of cancer in the lymph nodes and to where the cancer has spread (Metastases). In particular, the letter “T” stands for primary tumor:

“**T1a**” – tumor size ≤ 2 cm;

“**T1b**” – tumor size $> 2 \sim 3$ cm;

“**T2a**” – tumor size $> 3 \sim 5$ cm;

“**T2b**” – tumor size $> 5 \sim 7$ cm;

“**T3**” – tumor size > 7 cm and/or multiple tumor nodules in the same lobe;

“**T4**” – multiple tumor nodules (of any size) in the same lung but in a different lobe.

The letter “N” denotes the spread of the cancer to nearby lymph nodes:

“**N0**” – no regional lymph node metastasis;

“**N1**” – metastasis in ipsilateral peribronchial and/or ipsilateral hilar lymph nodes as well as intrapulmonary nodes including involvement by direct extension;

“**N2**” – metastasis in ipsilateral mediastinal and/or subcarinal lymph node(s);

“**N3**” – metastasis in contralateral mediastinal, contralateral hilar, ipsilateral or contralateral scalene, or supraclavicular lymph node(s).

The letter “M” describes distant metastasis:

“M1a” – malignant pleural or pericardial effusions and/or separate tumor nodules in the contralateral lung;

“M2b” – distant metastasis in extrathoracic organs.

Based on the seventh edition of the TNM classification system described above, we list the stage groupings for non-small-cell lung cancer in Supplementary Table 6. In order to have sufficient data amount for statistical analysis of each LUAD stage, in our work we combine stage Ia and Ib and label them as stage I, and do the same for Stages II, III, and IV.

Note 3: Statistical analysis and P-value

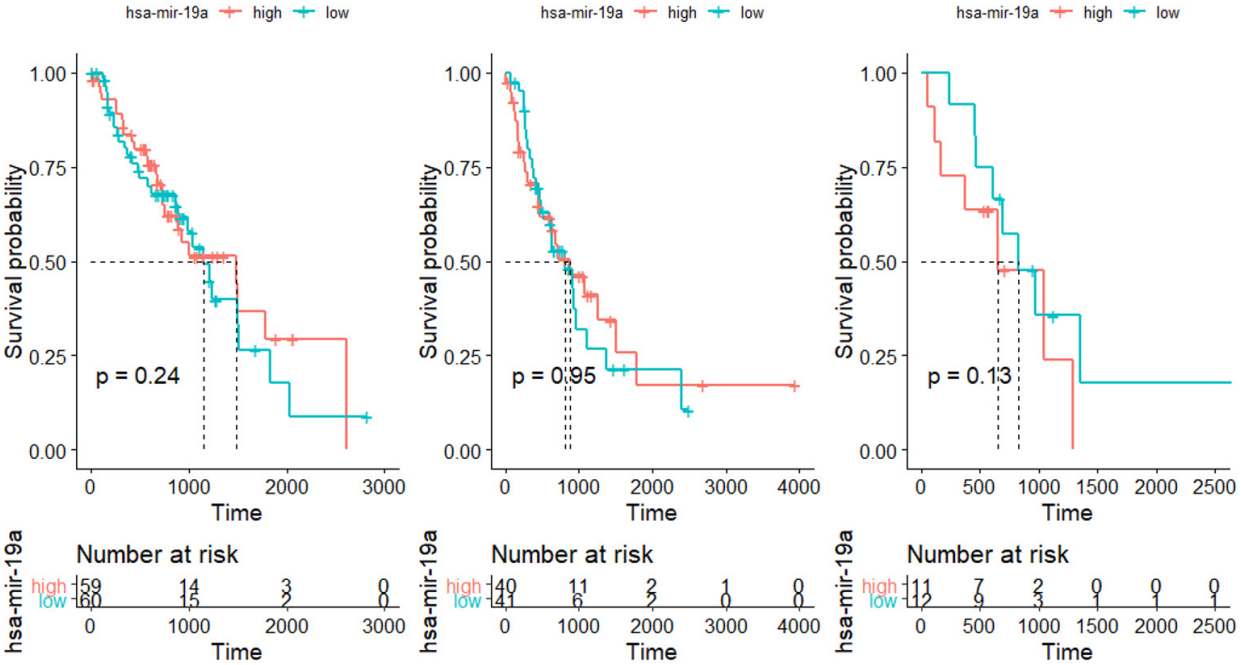
We have used four methods of statistical analysis involving hypothesis testing to calculate the P-values. The four methods are as follows.

- In our construction of the CeRNA network, RNAs are processed through a differentially expressed analysis, where the basic statistic used for significance assessment is the moderated t-statistic [1], which is computed for each contrast. This has the same interpretation as an ordinary t-statistic except that the standard errors have been moderated across genes, i.e., squeezed towards a common value, using a simple Bayesian model. This has the effect of borrowing information from the ensemble of genes to enhance inference about each individual gene. Moderated t-statistics lead to P-values in the same way that ordinary t-statistics do except that the degrees of freedom are increased, reflecting the greater reliability associated with the smoothed standard errors.
- The two-sample Wilcoxon test (also known as “Mann-Whitney” test) [2, 3] is applied to vectors of the RNA expression data, with the following notations for statistical significance: ns – $p > 0.05$; * – $p \leq 0.05$; ** – $p \leq 0.01$; *** – $p \leq 0.001$; **** – $p \leq 0.0001$.
- Gene enrichment analysis as described in MATERIALS AND METHODS and implemented in Figs. 3 and 5 in the main text.
- Log-rank tests as described at the end of subsection entitled “Framework of analysis” in MATERIALS AND METHODS in the main text.

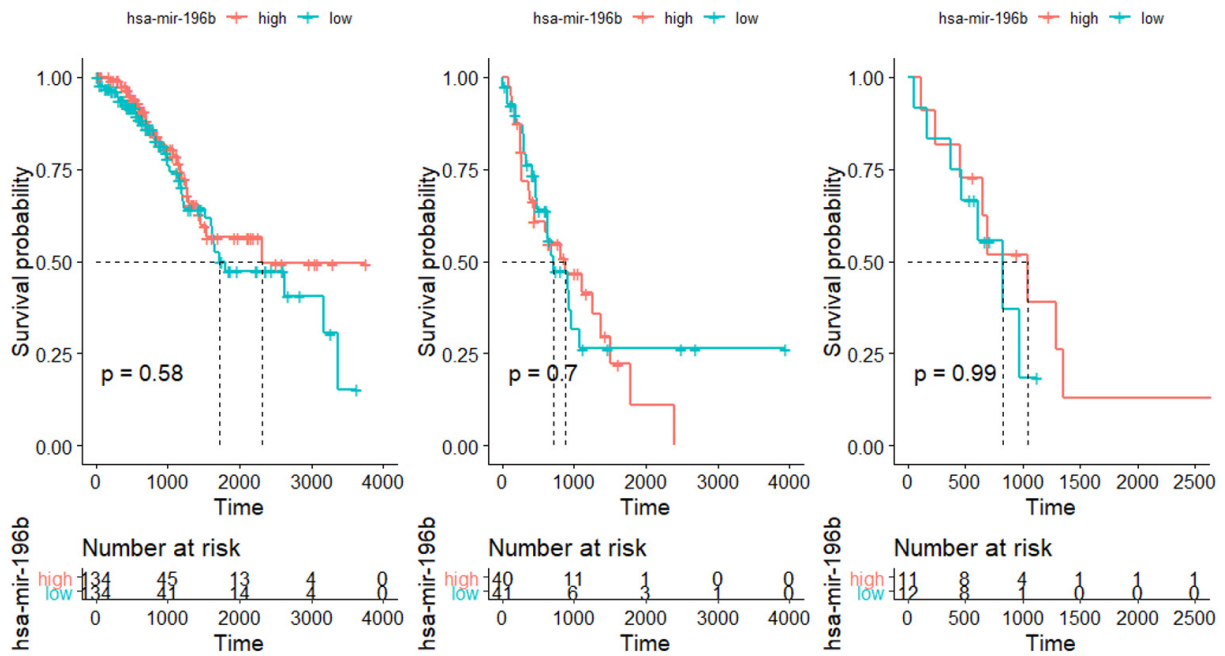
Note 4: Individual RNA expressions in the volcano map

Fold Change characterizes the relative expression level of interested samples to that of the control samples. For \log_2 Fold-Change = 1, the individual RNA expression level equals the group average of RNAs. The individual RNA behavior can be assessed from the volcano map of RNA expression. The volcano maps for the expressions of lncRNA, microRNA and mRNA at the four LUAD stages are shown in Supplementary Figs. 5-7.

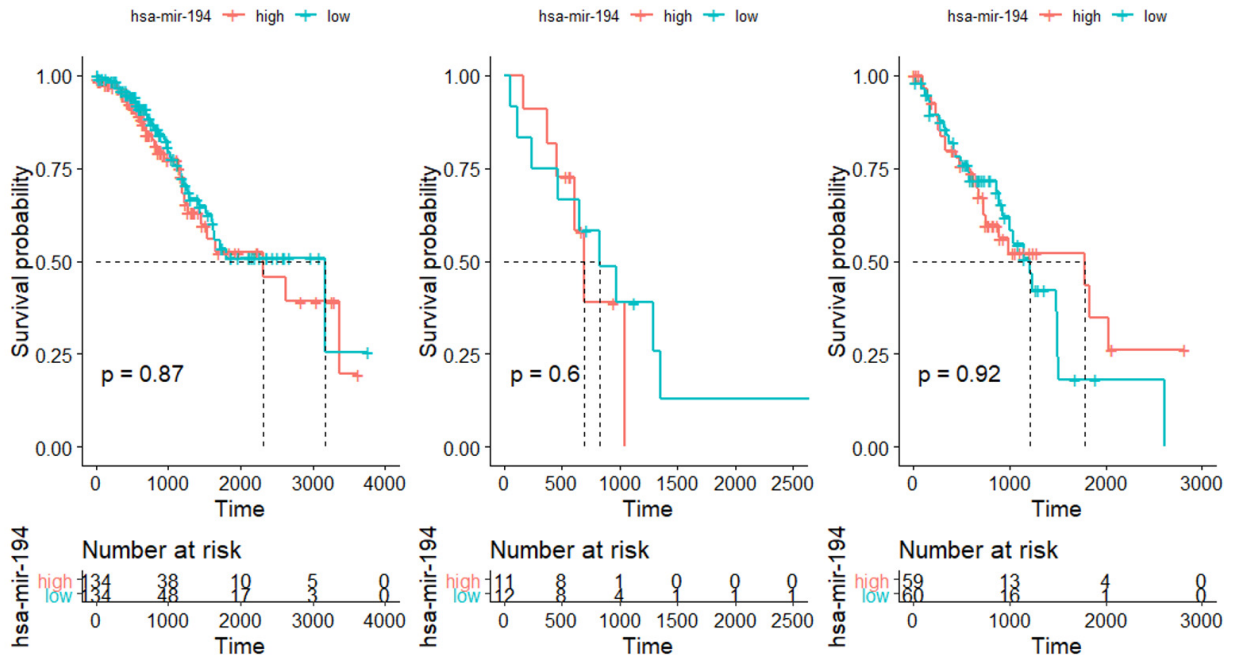
Supplementary Figures



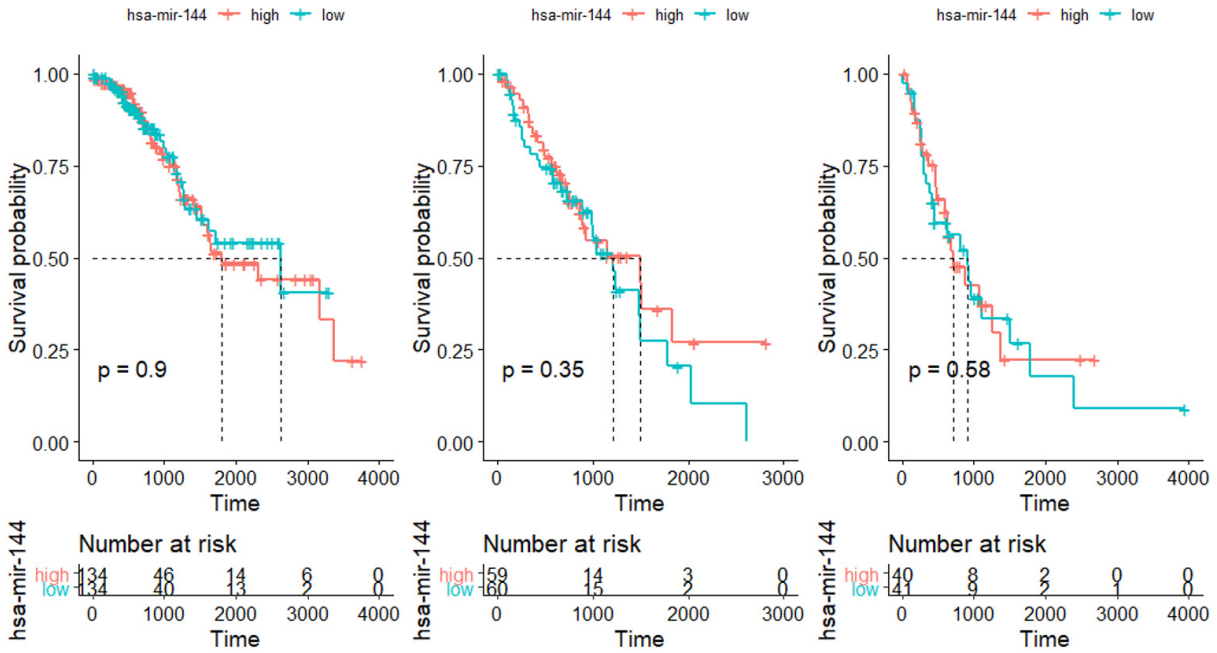
Supplementary Figure 1. Survival curves of microRNA hsa-mir-19a in stages II, III and IV. Log-rank tests are used to analysis of Kaplan-Meier survival curve.



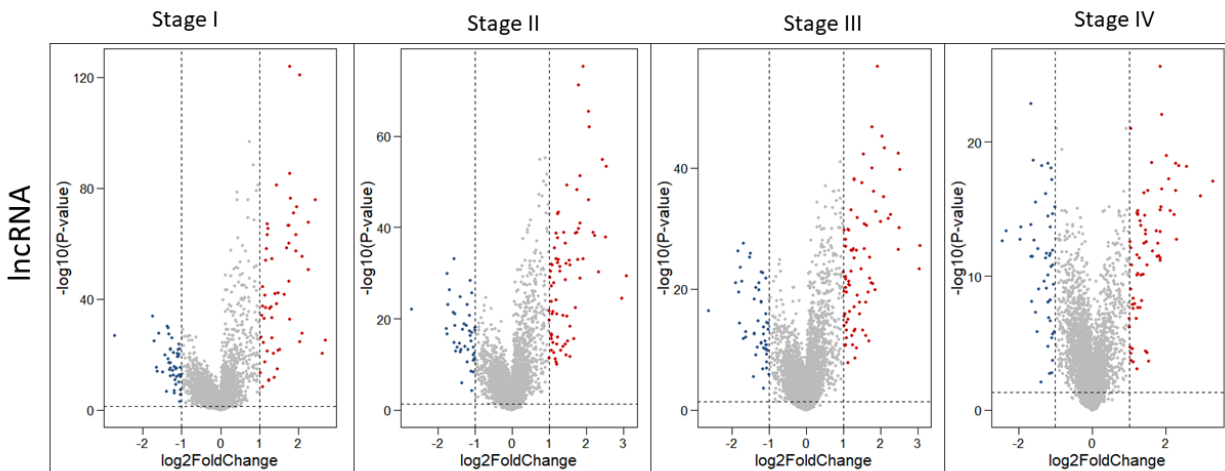
Supplementary Figure 2. Survival curves of microRNA hsa-mir-196b in stages I, III and IV. Log-rank tests are used to analysis of Kaplan-Meier survival curve. Log-rank tests are used to analysis of Kaplan-Meier survival curve.



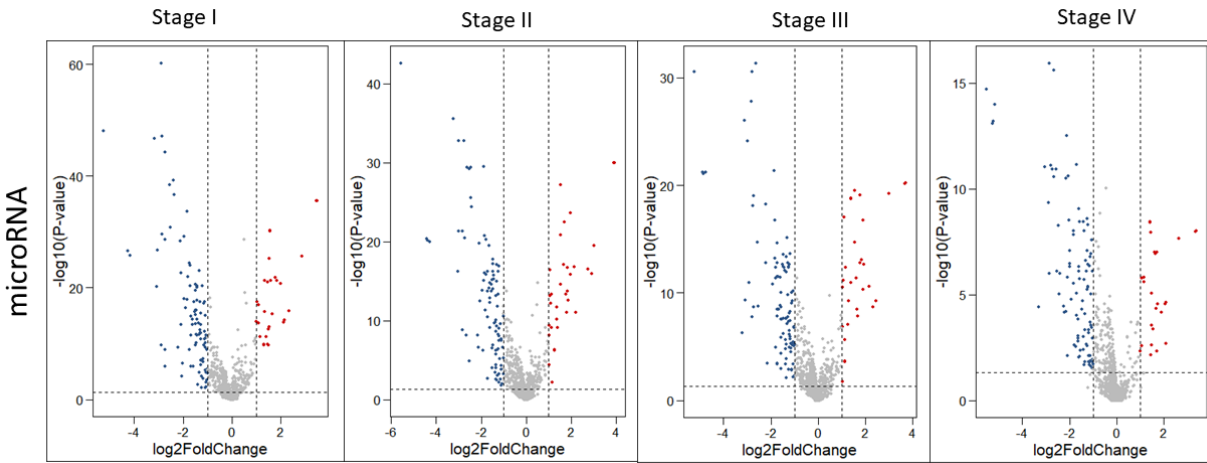
Supplementary Figure 3. Survival curves of microRNA hsa-mir-194 in stages I, II and IV. Log-rank tests are used to analysis of Kaplan-Meier survival curve.



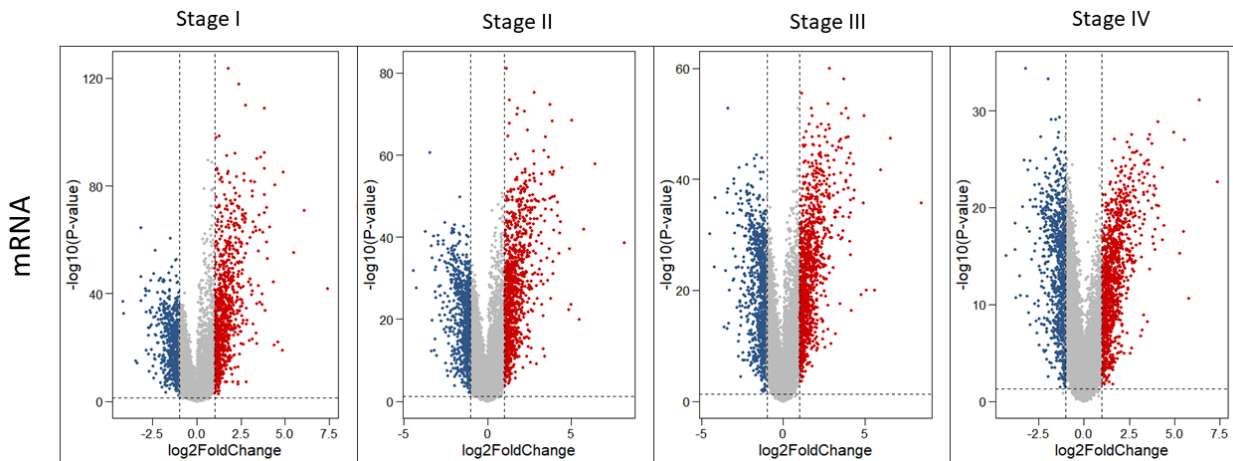
Supplementary Figure 4. Survival curves of microRNA hsa-mir-144 in stages I, II and III. Log-rank tests are used to analysis of Kaplan-Meier survival curve.



Supplementary Figure 5. Volcano map of the lncRNA expression level of four stages of LUAD samples. The x -axis is \log_2 Fold-Change (FC value), and the y -axis is the P-values from the differential expression analysis.



Supplementary Figure 6. Volcano map of the microRNA expression level of four stages of LUAD samples. The x -axis is \log_2 Fold-Change (FC value), and the y -axis is the P-values from the differential expression analysis.



Supplementary Figure 7. Volcano map of the mRNA expression level of four stages of LUAD samples. The x -axis is \log_2 Fold-Change (FC value), and the y -axis is the P-values from the differential expression analysis.

Supplementary Tables

Supplementary Table 1. Detailed parameter values of the reconstructed lncRNA-microRNA-mRNA CeRNA networks. The rows denoted as stages I, II, III, and IV correspond to the networks at the four stages of LUAD. The columns “total,” “dif,” and “same” represent the total number of nodes, the number of unique nodes, and the number of common nodes, respectively, for the networks at the four stages.

	Numbers of nodes and edges of CeRNA networks										
	lncRNA			microRNA			mRNA			all nodes	all edges
	total	dif	same	total	dif	same	total	dif	same		
Stage I	56	27		58	5		480	197		594	1633
Stage II	59	30		55	2		519	236		633	1603
Stage III	70	41	29	61	8	53	595	312	283	726	1952
Stage IV	65	36		57	4		618	335		740	1945

Supplementary Table 2. Clinical physiological data of the four LUAD stages matched with the lncRNA and mRNA expression data. Listed are the numbers of four LUAD samples and normal samples. The rows denoted as stages I, II, III, and IV represent the numbers of samples in the networks in the four stages of LUAD. The row “control” gives the number of normal samples.

	Distribution of clinical samples in gene expression data of LUAD		
	mRNA and lncRNA Data	microRNA Data	All samples
Control	59	46	85
Stage I	284	279	288
Stage II	122	122	124
Stage III	84	85	85
Stage IV	26	24	26
Total	575	556	608

Supplementary Table 3. Differentially expressed lncRNA, microRNA and mRNA profile data in the four stages of LUAD. The threshold of the absolute value of \log_2 FC (\log_2 fold change) for filtering the three types of RNAs is set to be one, ensuring that their P-values are less than 0.05 in the differential expression analysis. Fold changes greater than one correspond to up-regulated genes, while those less than minus one are associated with down-regulated genes. The rows Stages I, II, III, and IV represent the networks in the four stages of LUAD.

Differential expression (DE) gene number (\log_2 FC \geq |1|)
in the four stages of LUAD

	DE lncRNAs			DE microRNAs			DE mRNAs		
	total	up	down	total	up	down	total	up	down
Stage I	104	56	48	112	24	88	1657	966	691
Stage II	118	75	43	114	28	86	1890	1077	813
Stage III	131	80	51	118	26	92	2039	1159	880
Stage IV	119	69	50	110	22	88	2016	1169	847

Supplementary Table 4. The microRNA-mRNA and microRNA-lncRNA interactions relationships obtained by using the base of complementary pairing matching relationship data in the RNA sequences. The columns “total,” “up” and “down” represent the numbers of all, up-regulated and down-regulated RNAs in the four stages of LUAD. The columns miRNA-lncRNA and miRNA-mRNA list the numbers of the edges in the microRNA-lncRNA and microRNA-mRNA subnetworks, respectively, in the networks of four stages of LUAD.

Number of RNAs targeted by MicroRNAs and their relationships

	DE lncRNA			DE miRNA			DE mRNA			miRNA-lncRNA	miRNA-mRNA
	total	up	down	total	up	down	total	up	down		
Stage I	77	44	33	58	12	46	728	456	272	711	3726
Stage II	89	59	30	57	13	44	840	514	326	765	3982
Stage III	99	61	38	61	13	48	944	564	380	826	4792
Stage IV	93	53	40	58	12	46	904	556	348	808	4376

Supplementary Table 5. The Pearson correlation of the RNA expression data between microRNA-mRNA and microRNA-lncRNA expressions is calculated. The threshold of the correlation coefficient is set to be -0.1. The columns of miRNA-lncRNA and miRNA-mRNA list the numbers of edges for the microRNA-lncRNA and microRNA-mRNA subnetworks, respectively, in the networks of four stages of LUAD. The rows stage I, II, III, and IV correspond to the networks of the four stages of LUAD.

Number of RNAs targeted by MicroRNA and selected by introducing negative correlation between MicroRNAs and other RNAs (Cor<=-0.1)					
	DE lncRNA	DE miRNA	DE mRNA	miRNA-lncRNA	miRNA-mRNA
Stage I	56	58	480	202	1431
Stage II	59	55	519	206	1397
Stage III	70	61	595	276	1676
Stage IV	65	57	618	292	1653

Supplementary Table 6. Stage groupings for non-small-cell lung cancer

	Tumour (T)	Node (N)	Metastasis (M)
Ia	T1a or T1b	N0	M0
Ib	T2a	N0	M0
	T1a or T1b	N1	M0
IIa	T2a	N1	M0
	T2b	N0	M0
IIb	T2b	N1	M0
	T3	N0	M0
	T1 or T2	N2	M0
IIIa	T3	N1 or N2	M0
	T4	N0 or N1	M0
IIIb	T4	N2	M0
	any T	N3	M0
IV	any T	any N	M1a
	any T	any N	M1b

References

- [1] Phipson, B., Lee, S., Majewski, I. J., Alexander, W. S. & Smyth, G. K. Robust hyperparameter estimation protects against hypervariable genes and improves power to detect differential expression. *Ann. Appl. Stat.* **10**, 946 (2016).
- [2] Bauer, D. F. Constructing confidence sets using rank statistics. *J. Ame. Stat. Asso.* **67**, 687–690 (1972).
- [3] Hollander, M. & Wolfe, D. A. *Nonparametric Statistical Methods* (John Wiley & Sons, New York, 1973).