# Pan-cancer characterization of ncRNA synergistic competition uncovers potential carcinogenic biomarkers

Junpeng Zhang[1*], Lin Liu[2], Xuemei Wei[1], Chunwen Zhao[1], Sijing Li[1], Jiuyong Li[2], and Thuc Duy Le[2*]

[1] School of Engineering, Dali University, Dali, Yunnan 671003, China

[2] UniSA STEM, University of South Australia, Mawson Lakes, SA 5095, Australia

*To whom correspondence should be addressed. Email: zhangjunpeng411@gmail.com

Correspondence may also be addressed to Thuc Duy Le. Email: thuc.le@unisa.edu.au

## Abstract

Non-coding RNAs (ncRNAs) act as important modulators of gene expression and they have been confirmed to play critical roles in the physiology and development of malignant tumors. Understanding the synergism of multiple ncRNAs in competing endogenous RNA (ceRNA) regulation can provide important insights into the mechanisms of malignant tumors caused by ncRNA regulation. In this work, we present a framework, SCOM, for identifying ncRNA synergistic competition. We systematically construct the landscape of ncRNA synergistic competition across 31 malignant tumors, and reveal that malignant tumors tend to share hub ncRNAs rather than the ncRNA interactions involved in the synergistic competition. In addition, the synergistic competition ncRNAs (i.e. ncRNAs involved in the synergistic competition) are likely to be involved in drug resistance, contribute to distinguishing molecular subtypes of malignant tumors, and participate in immune regulation. Furthermore,

SCOM can help to infer ncRNA synergistic competition across malignant tumors and uncover potential diagnostic and prognostic biomarkers of malignant tumors. Altogether, the SCOM framework and the resulting web-based database SCOMdb (www.comblab.cn/SCOMdb/) serve as a useful resource for exploring ncRNA regulation and to accelerate the identification of carcinogenic biomarkers.

## Author summary

Abundant evidence reveals that ncRNAs are important modulators of gene expression, and it is important for us to understand the regulation of ncRNAs in malignant tumors. In this work, we hypothesize that ncRNAs acting as ceRNAs can crosstalk with each other in the form of synergistic competition. We name the hypothesis as the *ncRNA synergistic competition hypothesis*. To uncover potential carcinogenic biomarkers, we present a framework, SCOM, for identifying ncRNA synergistic competition in malignant tumors. From the perspective of synergistic competition, we have shown that the synergistic competition ncRNAs are likely to involve in drug resistance, contribute to distinguishing molecular subtypes of malignant tumors, and participate in immune regulation. In addition, the synergistic competition ncRNAs are potential diagnostic and prognostic biomarkers of malignant tumors. We believe that the proposed framework SCOM and the web-based database SCOMdb can contribute to the design of reliable diagnosis and treatment biomarkers for malignant tumors.

## Introduction

Malignant tumors are a type of complex human diseases characterized by uncontrolled growth and spread of malignant cells and tissue infiltration [1]. According to the latest global cancer statistics from International Agency for Research on Cancer (IARC), malignant tumors are the second leading cause of death worldwide, and severely threaten human life and health [2]. Depending on the type of malignant tumor cells, malignant tumors are mainly divided into two categories: cancer and sarcoma [3]. Compared with healthy cells, the structure and function of genes in tumor cells have changed. It is found that polygenic synergy is one of the key characteristics of malignant tumors [4], suggesting that malignant tumors are usually polygenic disorders and their pathogenesis is caused by multi-gene synergy.

Non-coding RNAs (ncRNAs) are a class of various RNA transcripts that usually cannot be translated into proteins. Ordinarily, ncRNAs are categorized into two types: regulatory ncRNAs and structural ncRNAs [5]. For example, microRNAs (miRNAs), long non-coding RNAs (lncRNAs), circular RNAs (circRNAs) and pseudogenes belong to the type of regulatory ncRNAs, while ribosome RNAs (rRNAs) and transfer RNAs (tRNAs) are considered as structural ncRNAs. Existing studies [6,7] have demonstrated that ncRNAs are closely associated with the occurrence and development of malignant tumors, which opens a promising avenue for the diagnosis and preventive treatment of malignant tumors.

As regulators, ncRNAs can modulate the expression of thousands of genes in the manner of trans-acting and cis-acting [8], thereby driving cell biological responses and fates [7]. In addition to the role of regulators, by soaking up miRNAs, ncRNAs can act as competing endogenous RNAs (ceRNAs) [9] which compete with the target messenger RNAs (mRNAs) of the soaked miRNAs. In the ceRNA context, miRNAs serve as mediators to connect non-coding and coding RNAs, resulting in the ceRNA network (consisting of ceRNAs and mRNAs), a more comprehensive regulatory network than the miRNA regulatory network [10]. So far, three types of ncRNAs, including lncRNAs, circRNAs and pseudogenes have been widely known to function as ceRNAs [11], and are regarded as potential carcinogenic biomarkers in clinical applications [12,13]. For example, lncRNA *H19* serves as a ceRNA for *LIN28* to promote breast cancer stem cell maintenance [14], circRNA *circTP63* functions as a ceRNA for *FOXM1* to accelerate lung squamous cell carcinoma progression [15], and pseudogene *PTENP1* acts as a ceRNA for *PTEN* in prostate cancer [16].

The ceRNA regulation of malignant tumors involves the interactions between multiple non-coding and coding genes. In the ceRNA network in which ncRNAs are involved, the competitive relationships between ncRNAs and target mRNAs are not one-to-one, but many-to-many. For example, previous studies [17–25] have demonstrated that the three ncRNAs, *H19*, *MALAT1* and *PVT1* synergistically compete with the four target genes, *KLF4*, *VEGFA*, *ZEB1* and *ZEB2*. In this work, we call such collective competition by a group of ncRNAs

against a group of mRNAs as “ncRNA synergistic competition”. It is observed that such complicated gene regulation carries out important tasks in physiological and pathological processes, and is considered to be a key component of cellular regulatory networks [7]. Furthermore, the entire cell system presents the characteristics of functional modularity, and each gene functional module (i.e. a group of genes interacting or crosstalking with each other frequently and not by chance) is responsible for specific biological functions in complex human diseases, including malignant tumors [26]. These findings indicate that ncRNAs, acting as ceRNAs, may synergistically rather than just independently play important roles in malignant tumors.

In this work, to uncover potential carcinogenic biomarkers, we present the SCOM (Synergistic COMpetition) framework to characterize ncRNA synergistic competition in Pan-cancer. SCOM first systematically predicts ncRNA-related ceRNA networks across 31 tumor types (containing 8001 tumor samples). Based on the proposed *ncRNA synergistic competition hypothesis*, SCOM applies three criteria, (i) significant synergistic competition for mRNAs, (ii) significant positive correlation, and (iii) significant sensitive correlation conditioning on the mRNAs synergistically competed against by the ncRNAs, to predict the ncRNA synergistic competition network (consisting of ncRNAs acting as ceRNAs) from the gene expression data of ncRNAs and target mRNAs and the predicted ncRNA-related ceRNA networks. By using SCOM, we provide a landscape of ncRNA synergistic competition in

Pan-cancer, contributing to the discovery of potential carcinogenic biomarkers. Ultimately, SCOM provides a framework for discovering the relationships between the synergistic competition ncRNAs (i.e. ncRNAs involved in the synergistic competition) and malignant tumors, and can help to elucidate ncRNA synergistic competition mechanisms in malignant tumors.

## Materials and methods

### **Synergistic COMpetition (SCOM)**

The proposed SCOM framework is based on the *ncRNA synergistic competition hypothesis* (**Fig 1a**). For soaking up miRNAs, a pool of ncRNA transcripts (acting as ceRNAs) competes with mRNAs through miRNA response elements (MREs). As a result, the competitive interactions between ncRNAs and mRNAs form a ceRNA network (consisting of ncRNAs acting as ceRNAs and target mRNAs). In the ceRNA network, we have observed that the competitive relationships between ncRNAs and mRNAs are not one-to-one but many-to-many, revealing synergistic competition of ncRNAs. Based on this observation, we hypothesize that ncRNAs acting as ceRNAs can crosstalk with each other in the form of synergistic competition (i.e. ncRNAs synergistically compete with mRNAs). We name the hypothesis as the *ncRNA synergistic competition hypothesis*.

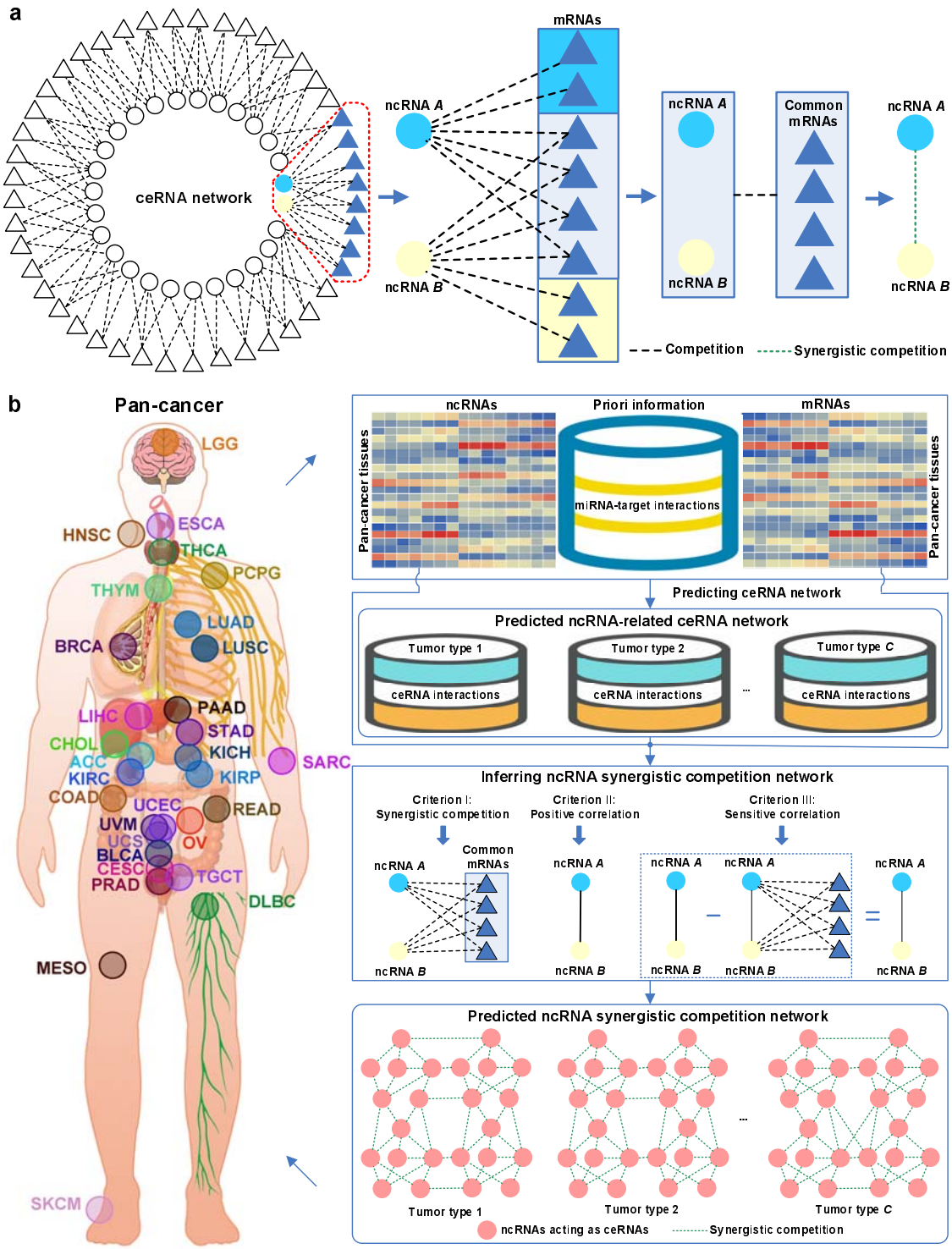



**Fig 1. Schematic illustration of SCOM.** (a) An illustration of the ncRNA synergistic competition hypothesis. ncRNA *A* and ncRNA *B* acting as ceRNAs collectively compete with mRNAs for soaking up miRNAs. In the context of ceRNA competition, the crosstalk between ncRNA *A* and ncRNA *B* is in the form of synergistic competition. (b) For each tumor type in Pan-cancer, SCOM firstly extends the SPONGE [33] method to predict ncRNA-related (including lncRNA- and pseudogene-related) ceRNA

network by incorporating gene (including miRNAs, lncRNAs, pseudogenes, and mRNAs) expression data in matched tumor tissues and priori information of miRNA-target (including miRNA-lncRNA, miRNA-pseudogene and miRNA-mRNA) interactions. By integrating gene (including lncRNAs and pseudogenes acting as ceRNAs, and target mRNAs) expression data in matched tumor tissues and predicted ncRNA-related ceRNA networks, SCOM utilise three criteria to infer ncRNA synergistic competition network for each tumor type in Pan-cancer. Based on the predicted ncRNA synergistic competition networks across 31 malignant tumors, SCOM further uncovers potential carcinogenic biomarkers.

The first step of SCOM (**Fig 1b**) is to predict ncRNA-related ceRNA network of each malignant tumor type. The matched ncRNA (including miRNA, lncRNA and pseudogene) and mRNA expression data in Pan-cancer (consisting of 8001 matched tumor samples) are obtained from The Cancer Genome Atlas (TCGA) [27] project. The information of miRNA targets is obtained by integrating five well-known databases, including miRTarBase v9.0 [28], TarBase v8.0 [29], LncBase v2.0 [30], NPInter 4.0 [31] and ENCORI [32]. By integrating gene (including ncRNAs and mRNAs) expression data and putative miRNA-target interactions, SCOM extends SPONGE [33] (a three-criterion evaluation method) to predict ncRNA-related ceRNA networks across 31 malignant tumors. The three-criterion evaluation method predicts ceRNA interactions by integrating sample-matched ncRNA (miRNA, lncRNA and pseudogene) and mRNA expression data, and putative miRNA-target interactions. We reason that if a candidate ncRNA-mRNA pair meets the three criteria (significant sharing of miRNAs, significantly positive correlation, and significantly sensitive correlation conditioning on shared miRNAs) with $p$-value $< 0.05$, the candidate ncRNA-mRNA pair is regarded as a ceRNA interaction. All of the predicted ceRNA interactions form

a large-scale ncRNA-related ceRNA network, with genes (including lncRNAs, pseudogenes and mRNAs) denoted as nodes and the ncRNA-mRNA competition relationships as undirected edges.

The second step of SCOM (**Fig 1b**) is to infer ncRNA synergistic competition network of each malignant tumor type. For a malignant tumor type, to predict the ncRNA synergistic competition network, we use the expression data of ncRNAs acting as ceRNAs and target mRNAs and the predicted ncRNA-related ceRNA networks. Following the *ncRNA synergistic competition hypothesis*, we consider a ncRNA-ncRNA pair having a synergistic competition relationship if the pair have (i) significant synergistic competition with mRNAs, (ii) significant positive correlation, and (iii) significant sensitive correlation conditioning on their competing mRNAs. The synergistic competition network of a malignant tumor type is then formed by integrating all of ncRNA synergistic competition relationships.

**Pan-cancer transcriptomics data**

After removing non-matching tumor samples between ncRNAs (including miRNAs, lncRNAs and pseudogenes) and mRNA expression data, the Pan-cancer transcriptomics data includes the expression data of 17,212 ncRNAs (including 894 miRNAs, 4366 lncRNAs, and 11,952 pseudogenes) and 19,068 mRNAs in 8001 matched tumor samples covering 31 tumor types (**S1 Data**), and the clinical information of 8001 matched tumor samples (**S2 Data**).

**Putative miRNA-target interactions**

Three types of putative miRNA-target interactions, including miRNA-mRNA, miRNA-lncRNA, and miRNA-pseudogene are used for predicting ncRNA-related ceRNA networks. The putative miRNA-mRNA interactions are acquired from miRTarBase v9.0 [28] and TarBase v8.0 [29], the putative miRNA-lncRNA interactions are obtained from LncBase v2.0 [30] and NPInter 4.0 [31], and the putative miRNA-pseudogene interactions are from ENCORI [32] (the previous version is starBase). In total, we have 762,540 miRNA-mRNA interactions, 225,063 miRNA-lncRNA interactions, and 91,619 miRNA-pseudogene interactions as priori information of miRNA-target interactions.

**Predicting ceRNA networks in Pan-cancer**

Given the gene expression data of miRNAs, lncRNAs, pseudogenes, and mRNAs, and miRNA-target interactions, three criteria, (i) significant sharing of miRNAs, (ii) significant positive correlation, and (iii) significant sensitive correlation conditioning on shared miRNAs, are used to predict ceRNA network (see **Supplementary Methods** in **S1 File**).

In this work, a candidate ncRNA-mRNA pair meeting the above three criteria with $p$-value $< 0.05$ is regarded as a ceRNA interaction. For each of the 31 tumor types, we integrate all of ceRNA interactions to form the ceRNA network of the tumor type.

**Constructing ncRNA synergistic competition networks in Pan-cancer**

Given the gene expression data of lncRNAs and pseudogenes acting as ceRNAs, and miRNA targets mRNAs and the predicted ncRNA-related ceRNA networks, three criteria, (i) significant synergistic competition for mRNAs, (ii) significant positive correlation, and (iii) significant sensitive correlation conditioning on synergistically competed mRNAs are used to construct ncRNA synergistic competition network.

The first criterion uses a hypergeometric test to assess the significance of synergistic competition of ncRNA *i* and ncRNA *j* for mRNAs. The significance *p*-value of the test is computed as:

$$p_{ij} = 1 - \sum_{x_{ij}=0}^{L_{ij}-1} \frac{\binom{M_{ij}}{x_{ij}}\binom{N_{ij}-M_{ij}}{K_{ij}-x_{ij}}}{\binom{N_{ij}}{K_{ij}}} \quad (1)$$

where $N_{ij}$ is the number of all mRNAs in a Pan-cancer dataset, $M_{ij}$ and $K_{ij}$ represent the total numbers of mRNAs competing with ncRNA *i* and ncRNA *j* respectively, and $L_{ij}$ is the number of mRNAs synergistically competed by ncRNA *i* and ncRNA *j*.

The second criterion requires that ncRNA *i* and ncRNA *j* are positively correlated. The Pearson [34] method is also utilized to evaluate whether the expression level of ncRNA *i* and the expression level of ncRNA *j* are positively correlated at a significant level (e.g. 0.05). The correlation $cor_{ij}$ between ncRNA *i* and ncRNA *j* is computed as:

$$cor_{ij} = cor(p,q) = \frac{\sum_{k=1}^{s}(p_k - \overline{p})(q_k - \overline{q})}{\sqrt{\sum_{k=1}^{s}(p_k - \overline{p})^2}\sqrt{\sum_{k=1}^{s}(q_k - \overline{q})^2}} \quad (2)$$

where $p = (p_1, p_2, ..., p_s)$ and $q = (q_1, q_2, ..., q_s)$ denote the expression level of ncRNA *i* and ncRNA *j* respectively, $\overline{p}$ and $\overline{q}$ represent the average expression level of ncRNA *i* and ncRNA *j* respectively, and *s* is the number of samples for each tumor type. Accordingly, the significance *p*-value $pc_{ij}$ is calculated as:

$$pc_{ij} = 2pt\left(cor_{ij}\sqrt{\frac{s-2}{1-(cor_{ij})^2}}\right) \quad (3)$$

where the *pt* function is used to calculate the probability of $cor_{ij}\sqrt{\frac{s-2}{1-(cor_{ij})^2}}$.

The third criterion is used to evaluate the influence of synergistically competed mRNAs on ncRNA *i* and ncRNA *j*. The sensitive correlation [35] is also employed to calculate the influence of the mRNAs synergistically competed by ncRNA *i* and ncRNA *j*, and the null model [33] is also applied to evaluate the significance of this influence. The sensitive correlation $sc_{ij}$ is calculated as:

$$sc_{ij} = cor_{ij} - pcor_{ij} \quad (4)$$

where $pcor_{ij}$ is the partial correlation between ncRNA *i* and ncRNA *j*, i.e. the correlation conditioning on the mRNAs synergistically competed by the ncRNAs, which is the correlation between ncRNA *i* and ncRNA *j* when the influence of the synergistically competed mRNAs is eliminated. $pcor_{ij}$ is calculated as:

$$pcor_{ij} = cor(p,q \mid M) = cor(p,q \mid (M_1,M_2,...,M_n))$$

$$= \frac{cor(p,q \mid (M_1,M_2,...,M_{n-1})) - cor(p,M_n \mid (M_1,M_2,...,M_{n-1}))cor(q,M_n \mid (M_1,M_2,...,M_{n-1}))}{\sqrt{1-cor(p,M_n \mid (M_1,M_2,...,M_{n-1}))^2}\sqrt{1-cor(q,M_n \mid (M_1,M_2,...,M_{n-1}))^2}} \quad (5)$$

where $p$, $q$ and $M$ denote the expression level of ncRNA $i$, ncRNA $j$, and the synergistically competed $n$ mRNAs respectively, $cor(p,q \mid (M_1,M_2,...,M_n))$ denotes the partial correlation between $p$ and $q$ conditioning on $(M_1,M_2,...,M_n)$, $cor(p,q \mid (M_1,M_2,...,M_{n-1}))$ represents the partial correlation between $p$ and $q$ conditioning on $(M_1,M_2,...,M_{n-1})$, $cor(p,M_n \mid (M_1,M_2,...,M_{n-1}))$ is the partial correlation between $p$ and $M_n$ conditioning on $(M_1,M_2,...,M_{n-1})$, $cor(q,M_n \mid (M_1,M_2,...,M_{n-1}))$ is the partial correlation between $q$ and $M_n$ conditioning on $(M_1,M_2,...,M_{n-1})$.

To evaluate the significance of $sc_{ij}$, the null model hypothesizes that the synergistically competed mRNAs do not affect the correlation between ncRNA $i$ and ncRNA $j$, i.e. the sensitive correlation (the difference between $cor_{ij}$ and $pcor_{ij}$) between ncRNA $i$ and ncRNA $j$ is 0. The number of datasets sampled is set to 1E+03 for the null model, and the pre-computed covariance matrices in the R package SPONGE [33] are used to build the null model. Based on the constructed null model, the significance $p$-value of $sc_{ij}$ is calculated.

In this study, a candidate ncRNA-ncRNA pair meeting the three criteria for synergistic competition with $p$-value < 0.05 is considered as a ncRNA synergistic competition relationship. For each tumor type, we combine all of the ncRNA synergistic competition relationships to create the ncRNA synergistic competition network for the tumor type.

**Network topological analysis**

Network topological features yield useful insights into biological networks, e.g. degree, path and density. If the node degree of a biological network obeys a power law distribution, the biological network is regarded as a scale-free network. The path and density features are used to measure the tightness of a biological network. Compared with the random networks, a shorter characteristic path length or higher density of a biological network means that the biological network is more likely to be a small-world network [36,37].

The topological features (degree, path and density) of ncRNA-related ceRNA networks and ncRNA synergistic competition networks are analyzed using the R package igraph [38], a useful tool for network analysis and visualization. The ncRNA-related ceRNA networks and ncRNA synergistic competition networks are undirected networks, and the degree of a node is the number of edges connected with the node. The Kolmogorov-Smirnov (KS) test [39] is used to evaluate whether the node degree of a network follows a power law distribution. If the *p*-value of the KS test is smaller than a cutoff (e.g. 0.05), the node degree of a ncRNA-related ceRNA network or a ncRNA synergistic competition network does not follow a power law distribution, then these indicate that the ncRNA-related ceRNA network or the ncRNA synergistic competition network is not a scale-free network. To determine whether the ncRNA-related ceRNA networks and the ncRNA synergistic competition networks are small-world networks, we generate 100 random networks (the same number of

nodes and edges with real network) for each ncRNA-related ceRNA network or ncRNA synergistic competition network. The Student's *t*-test [40] is applied to evaluate whether the characteristic path length (or density) of a ncRNA-related ceRNA network or a ncRNA synergistic competition network is smaller (or higher) than that of its corresponding random networks at a significance level. For a ncRNA-related ceRNA network or a ncRNA synergistic competition network, if the *p*-value of the Student's *t*-test is smaller than a cutoff (e.g. 0.05), the characteristic path length (or density) of a ncRNA-related ceRNA network or a ncRNA synergistic competition network is significantly smaller (or higher) than that of its corresponding random networks, then the network is a small-world network.

**Network and hub similarity calculation**

In terms of ncRNA synergistic competition networks or hub ncRNAs, the network or hub similarity between tumor *c* and tumor *d* is computed as follows.

$$sim_{cd} = \frac{|SCR_c \cap SCR_d|}{min(|SCR_c|, |SCR_d|)} \tag{6}$$

where $SCR_c$ and $SCR_d$ denote the identified ncRNA synergistic competition regulation (networks or hub ncRNAs) in tumor *c* and tumor *d*, respectively, $|SCR_c \cap SCR_d|$ is the number of common ncRNA synergistic competition relationships or hub ncRNAs between $SCR_c$ and $SCR_d$, and $min(|SCR_c|, |SCR_d|)$ denotes the smaller number of ncRNA synergistic competition relationships or hub ncRNAs between $SCR_c$ and $SCR_d$.

**Hub ncRNA identification**

Hub ncRNAs are defined as highly connected nodes in a ncRNA synergistic competition network. Here, we use the cumulative probability of Poisson distribution to decide whether a ncRNA is a hub ncRNA:

$$p(x \geq k) = 1 - \sum_{i=0}^{k-1} \frac{\lambda^i e^{-\lambda}}{i!} \tag{7}$$

where $\lambda = np$, $p = \frac{m}{A_n^2}$, $n$ is the number of ncRNAs, $m$ is the number of ncRNA synergistic competition relationships in a ncRNA synergistic competition network, and $A_n^2$ is the number of all possible ncRNA synergistic competition pairs. Smaller *p*-value of a ncRNA indicates that the ncRNA is more likely to be a hub ncRNA. In this work, the cutoff of the *p*-value is set to 0.05.

**Uncovering conserved and rewired interactions and hubs**

The ncRNA synergistic competition interaction or hub ncRNA existing in at least two malignant tumors is defined as a conserved ncRNA synergistic competition interaction or hub ncRNA, whereas the ncRNA synergistic competition interaction or hub ncRNA existing in only one malignant tumor is defined as a rewired ncRNA synergistic competition interaction or hub ncRNA. In the case of the identified ncRNA synergistic competition networks and hub ncRNAs across 31 malignant tumors, we can uncover conserved and rewired ncRNA synergistic competition interactions and hub ncRNAs.

## Multi-class classification analysis

To understand the classification performance of the discovered conserved ncRNA synergistic competition interactions and hub ncRNAs, we train a classification model to evaluate the discriminating ability of the features in classifying tumor types, and also train a model to classify BRCA subtypes (LumA, LumB, Her2, Basal and Normal) to understand the ability of the features in discriminating cancer subtypes. We carry out multi-class classification analysis by using the Binary Relevance (BR) [41] strategy in the R package utiml [42] to perform multi-class classification analysis. In the BR strategy, the Support Vector Machine (SVM) [43] implemented in the R package e1071 [44] is selected as the base classifier to create the multi-class model. We use 22 metrics (accuracy, average-precision, clp, coverage, F1, hamming-loss, macro-AUC, macro-F1, macro-precision, macro-recall, margin-loss, micro-AUC, micro-F1, micro-precision, micro-recall, mlp, one-error, precision, ranking-loss, recall, subset-accuracy and wlp as described in **S1 File**) and conduct 10-fold cross-validation to evaluate the performance of ncRNAs in conserved ncRNA synergistic competitions and hub ncRNAs. Also, we apply a commonly used baseline method [45] for comparison in the evaluation of the classification performance.

## Survival analysis

To compute the risk score of each malignant tumor sample, the multivariate Cox model [46] in the R package survival [47] is fitted by using ncRNAs in conserved ncRNA synergistic

competition and hub ncRNAs as features. According to the risk scores of each malignant tumor sample, the top one-third of the malignant tumor samples are put in the high risk group and the remaining two thirds are put in the low risk group. In terms of the gene expression levels of the ncRNAs, the Log-rank test is applied to assess the distinction of the feature ncRNAs between the high and the low risk groups or between different molecular subtypes. In addition, the proportional hazard ratio (HR) between the high and the low risk groups or between different molecular subtypes is also computed.

**Enrichment analysis**

To understand the biological significance of hub ncRNAs across the 31 malignant tumors, we conduct enrichment analysis in three categories, disease, cell marker and cancer hallmark using LncSEA [48]. The hub ncRNAs are enriched in significant disease, cell marker and cancer hallmark terms with an adjusted *p*-value (adjusted by Benjamini-Hochberg method) cutoff of 0.05. Moreover, LncSEA is also used to uncover whether the synergistic competition ncRNAs are involved in drug resistance or not. The synergistic competition ncRNAs are significantly enriched in drug terms with an adjusted *p*-value (adjusted by Benjamini-Hochberg method) less than 0.05.

**Molecular subtyping of malignant tumors**

Within the same type of malignant tumor, malignant tumors have great phenotypic differences and genetic heterogeneity. Because of this, molecular subtyping is usually applied

to divide malignant tumors into subtypes with distinct molecular and clinical characteristics. To assess the performance of the synergistic competition ncRNAs in molecular subtyping, we use an ensemble method implemented in the CancerSubtypes [49] R package (combining the Similarity Network Fusion (SNF) [50] and Consensus Clustering (CC) [51] methods) to generate malignant tumor subtypes. For each malignant tumor, the number of molecular subtypes is empirically set according to the previous molecular subtyping studies (**S3 Data**).

### Immune regulation analysis

Numerous ncRNAs have been shown to play important roles in immune regulation. For the immune regulation analysis, we obtain the significant immune-related ncRNA-pathway and ncRNA-cell pairs from ImmReg [52]. Based on the significance of the enrichment of ncRNAs with immune pathways and the significance of the correlation between ncRNAs and immune cells, we further determine which synergistic competition ncRNAs are significantly enriched in immune pathways and correlated with immune cells in each malignant tumor.

## Results

### The ceRNA regulation landscape in Pan-cancer

Using the transcriptomics data in TCGA, we apply SCOM for detecting two types of ceRNA regulation landscape across the 31 malignant tumors. In the ceRNA regulation landscape, the two types of ceRNA networks (lncRNA- and pseudogene-related ceRNA networks) have different numbers of ceRNA interactions across the 31 malignant tumors. We have found that

the number of ceRNA interactions of the pseudogene-related ceRNA networks is significantly larger than that of the lncRNA-related ceRNA networks (**Fig 2a**, *p*-value = 1.84E-13 with paired Student's *t*-test). Network topological analysis reveals that 5 out of the 31 (16.13%) lncRNA-related ceRNA networks and 16 out of the 31 (51.61%) pseudogene-related ceRNA networks follow a power law distribution (**Fig 2b**). After merging the lncRNA and pseudogene related ceRNA networks of each tumor, 14 out of the 31 (45.16%) merged ncRNA-related ceRNA networks follow a power law distribution (**Fig S1** in **S1 File**). Moreover, 7 out of the 31 (22.58%) lncRNA-related ceRNA networks and none of the pseudogene-related ceRNA networks have shorter characteristic path lengths than their corresponding random networks, and 30 out of the 31 (96.77%) lncRNA (and pseudogene related) ceRNA networks have higher densities than their corresponding random networks (**Fig 2b**). In the merged ncRNA-related ceRNA networks of each tumor, we have found that none of the ncRNA-related ceRNA networks have shorter characteristic path lengths than their corresponding random networks, and all of the ncRNA-related ceRNA networks have higher densities than their corresponding random networks (**Fig S1** in **S1 File**). These observations indicate that less than or nearly half of the ceRNA networks across the 31 malignant tumors tend to be scale-free networks, and most or all of the ceRNA networks across the 31 malignant tumors are likely to be small-world networks with higher densities than their corresponding random networks.

With the lncRNA and pseudogene related ceRNA networks, the similarity intervals between the 31 malignant tumors are [0.13 0.52] and [0.10 0.44], respectively (**Fig 2c**). Moreover, in terms of the ncRNA (including both lncRNA and pseudogene) related ceRNA networks, the similarity interval between the 31 malignant tumors is [0.12 0.47] (**Fig S2** in **S1 File**). This result shows that the malignant tumors may have common ncRNA-related ceRNA regulation. We have also discovered that each tumor ceRNA network varies in similarity to other tumor ceRNA networks, revealing known and hidden relationships among these malignant tumors. In addition, we have observed that ~16.45% lncRNA-related ceRNA regulations and ~28.58% pseudogene-related ceRNA regulations only exist in one malignant tumor (**Fig 2d**). Overall, ~24.80% ncRNA-related ceRNA regulations are occurred in one malignant tumor (**Fig S3** in **S1 File**). For malignant tumors, the percentages of the rewired ncRNA-related ceRNA regulations are different, demonstrating different degrees of heterogeneity across malignant tumors (**Fig S4** in **S1 File**). Particularly, 106 (less than 0.01%) ncRNA-related ceRNA regulations are conserved in the 31 malignant tumors (**Fig 2e**), indicating a low conservation of ceRNA regulation in Pan-cancer. By conducting the multi-class classification and survival analysis, we have shown that the conserved ncRNA-related ceRNA regulations in the 31 malignant tumors have better performance than the baseline method in classifying 31 malignant tumors, and can distinguish the high and the low risk groups of each malignant tumor at a significant level (Log-rank $p$-value $< 0.05$ and HR $> 2$)

(**Fig S5** in **S1 File**). This result reveals that the conserved ncRNA-related ceRNA regulations may act as a common core of ceRNA regulation across malignant tumors.

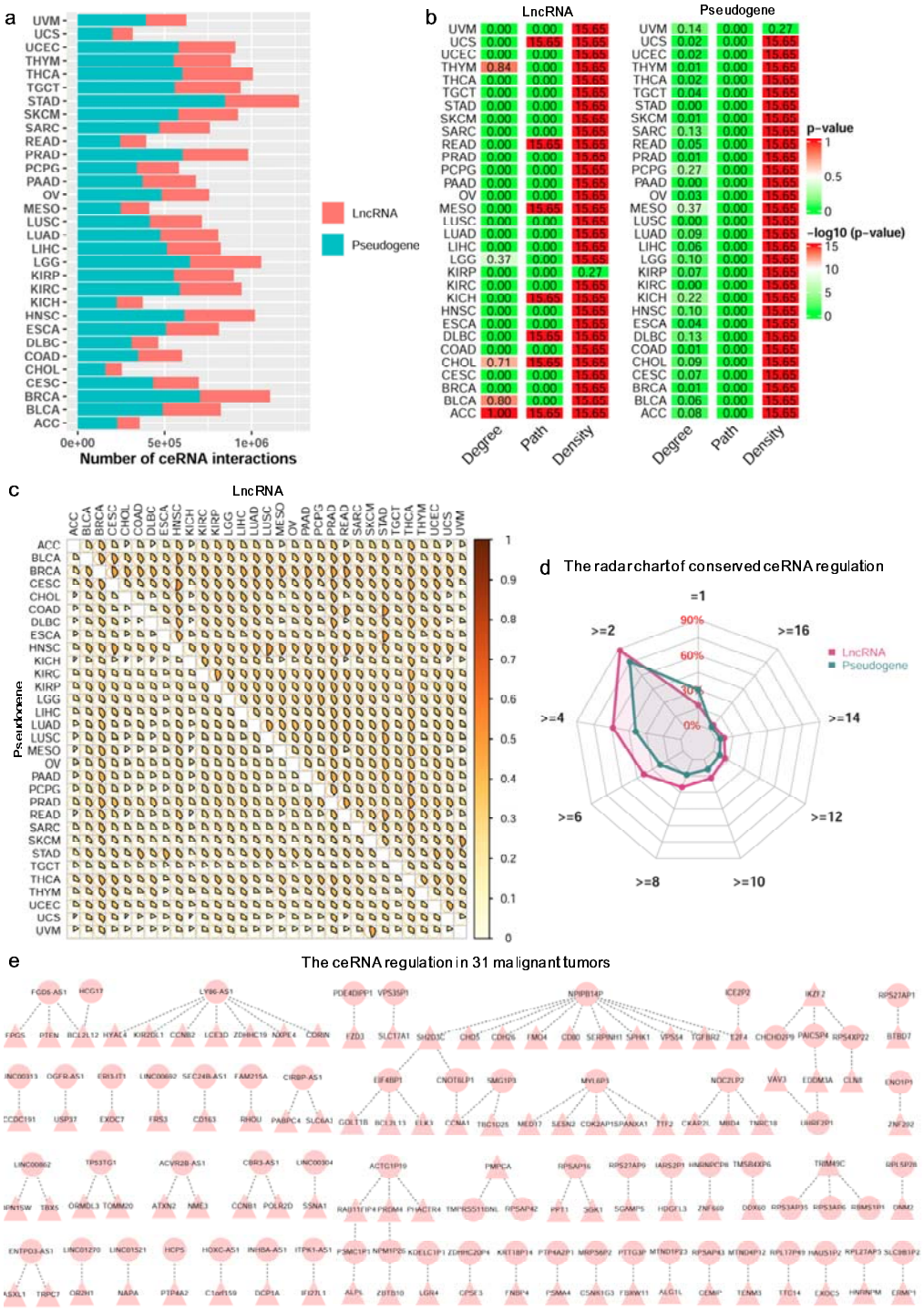



**Fig 2. The ceRNA regulation landscape across malignant tumors.** (a) The number of ncRNA-related ceRNA interactions in each malignant tumor. (b) The topological characteristics of ncRNA-related ceRNA

networks for the 31 malignant tumors. The Degree column is the *p*-value of fitting power-law distribution. The *p*-values equal to or more than 0.05 indicate that the degree of ncRNA-related ceRNA networks obeys the power-law distribution. The Path and Density columns represent the –log10(*p*-value) of Student's *t*-test, and the *p*-values less than 0.05 display that the characteristic path lengths (or densities) of ncRNA-related ceRNA networks are significantly shorter (or higher) than those of their corresponding random networks. (c) The similarity matrix shows the similarity between each pair of lncRNA (upper triangle) and pseudogene (lower triangle) related ceRNA networks across 31 malignant tumors. (d) The radar chart shows the percentage of ceRNA regulations predicted in different numbers of malignant tumors. (e) The conserved ceRNA regulation predicted in 31 malignant tumors.

## The ncRNA synergistic competition landscape in Pan-cancer

Taking advantage of the transcriptomics data in TCGA and the predicted ncRNA-related ceRNA interactions, we utilise SCOM for uncovering ncRNA (including lncRNAs and pseudogenes) synergistic competition landscape across 31 malignant tumors. Similar to the number of ceRNA interactions, the number of ncRNA synergistic competition interactions (including lncRNA-lncRNA, lncRNA-pseudogene and pseudogene-pseudogene synergistic competition pairs) also varies across the 31 malignant tumors. The largest and least numbers of ncRNA synergistic competition interactions are inferred in Stomach adenocarcinoma (STAD) and Cholangiocarcinoma (CHOL), respectively (**Fig 3a**). This result may be explained in part by the different number of ceRNA interactions predicted in the 31 malignant tumors. Network topological analysis shows that 27 out of the 31 (87.10%) ncRNA synergistic competition networks follow a power law distribution (**Fig 3b**). Furthermore, 19 out of the 31 (61.29%) ncRNA synergistic competition networks have shorter characteristic path lengths than their corresponding random networks, and all of the

ncRNA synergistic competition networks have higher densities than their corresponding random networks (**Fig 3b**). These findings indicate that most of the ncRNA synergistic competition networks across the 31 malignant tumors are scale-free networks, and small-world networks with shorter characteristic path lengths and higher densities than their corresponding random networks.

With the ncRNA synergistic competition networks, the similarity among the 31 malignant tumors is less than 6% (**Fig 3c**). This low similarity indicates that the malignant tumors may share a small portion of ncRNA synergistic competition with each other. In addition, ~94.06% ncRNA synergistic competition interactions only occur in one malignant tumor, and none of the ncRNA synergistic competition interactions exist in more than six malignant tumors (**Fig 3d**). Specifically, we have observed that the percentage of the conserved ncRNA synergistic competition interactions in at least two malignant tumors is ~5.94%. This result reveals that the ncRNA synergistic competition interactions of different malignant tumors tend to be rewired in their respective tumor microenvironments. To understand whether the ncRNAs of the conserved synergistic competition in at least two malignant tumors are biologically meaningful, we further perform enrichment analysis in three categories including diseases, cell markers and cancer hallmarks. The functional enrichment analysis has revealed that the ncRNAs of the conserved synergistic competition in at least two malignant tumors are significantly enriched in diseases (e.g. Cancer), cell markers (e.g. Exhausted CD8+ T cell),

and cancer hallmarks (e.g. Epithelial-mesenchymal transition) common to multiple malignant tumors (**S4 Data**). Moreover, the percentage of the rewired ncRNA synergistic competition for each malignant tumor is all more than 80%, demonstrating a great heterogeneity across malignant tumors in terms of ncRNA synergistic competition (**Fig 3e**).

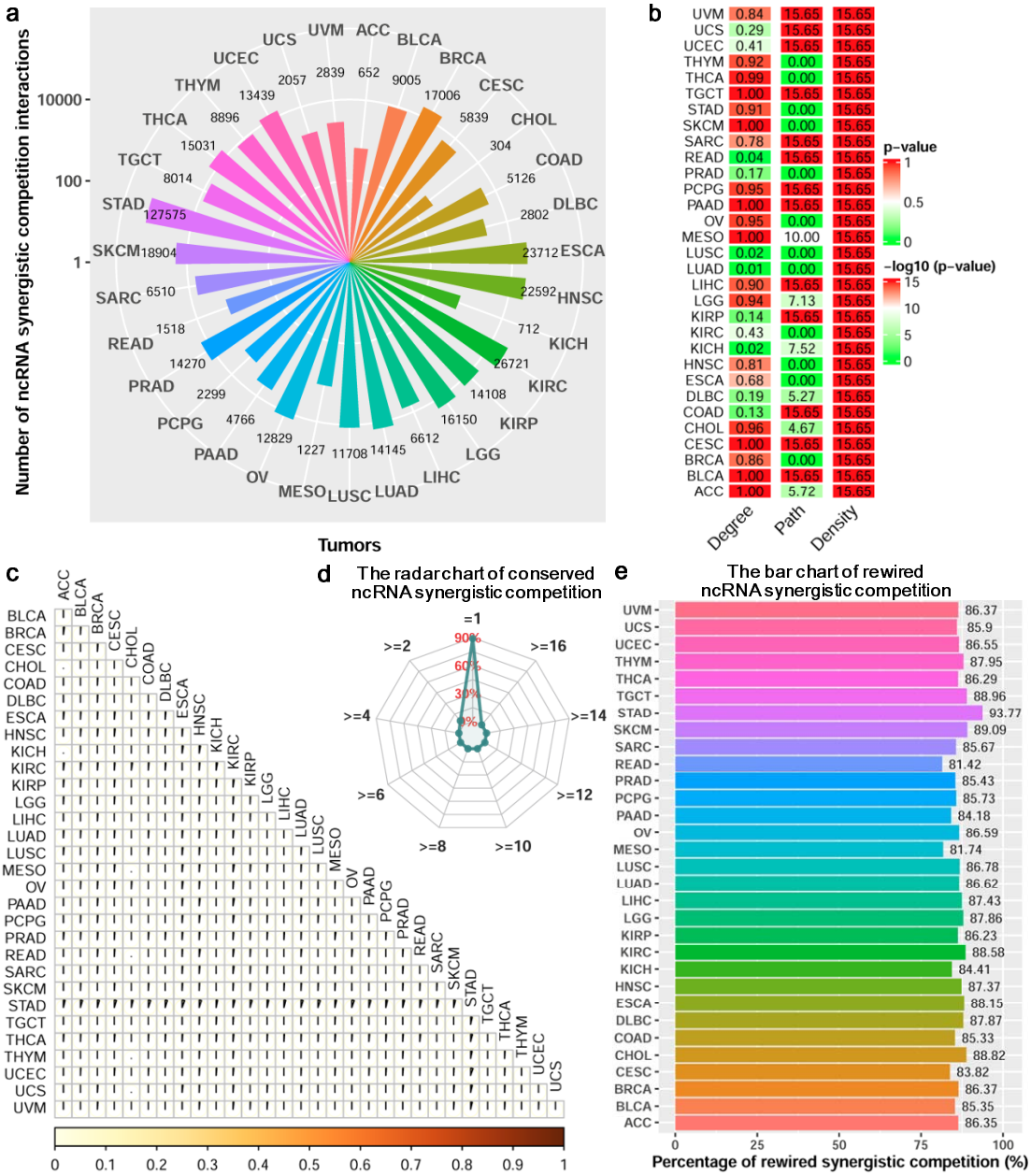



**Fig 3. Network analysis of ncRNA synergistic competition across malignant tumors.** (a) The number of ncRNA synergistic competition interactions in each malignant tumor. (b) The topological characteristics

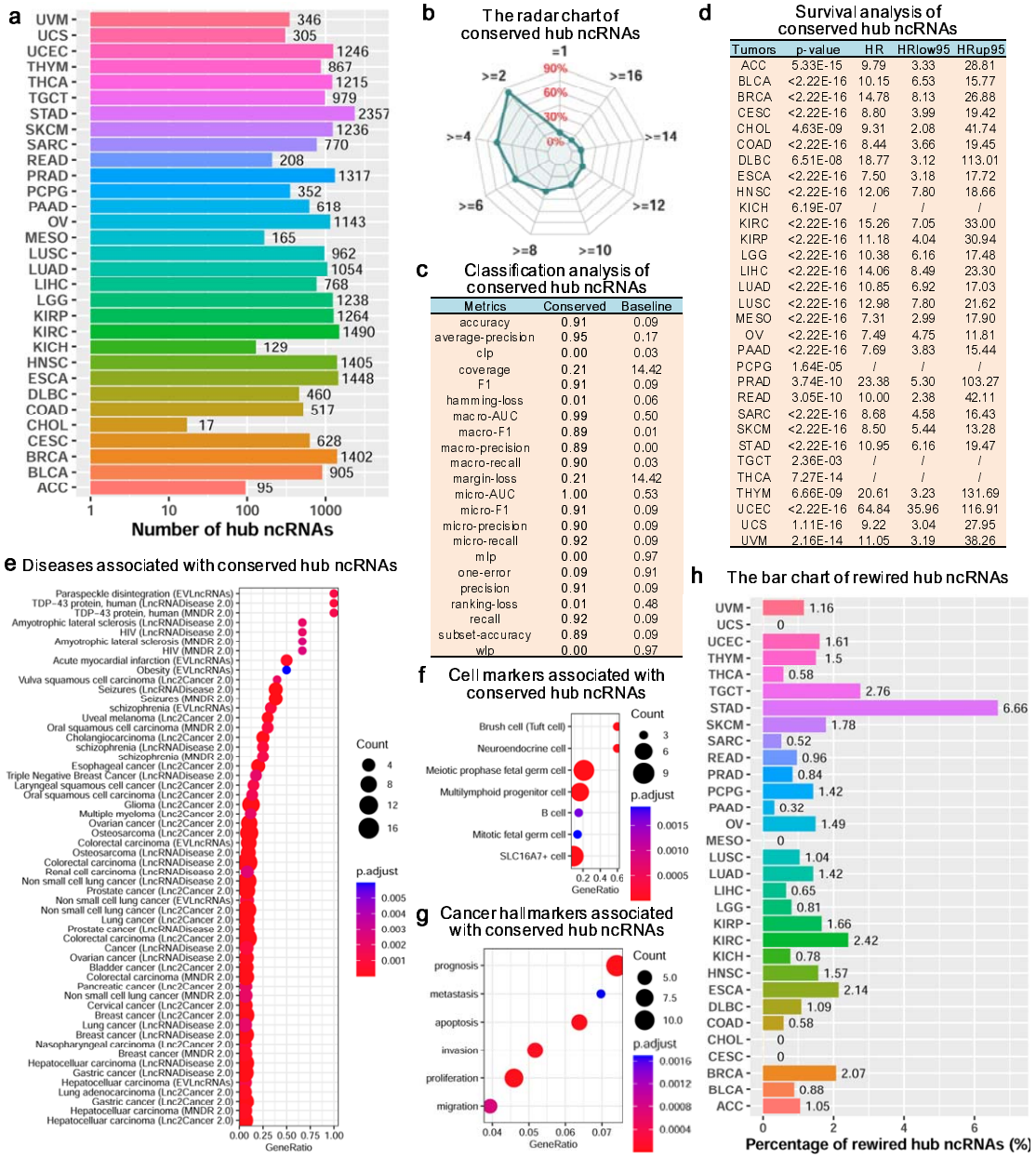
of ncRNA synergistic competition networks for the 31 malignant tumors. The Degree column is the *p*-value of fitting power-law distribution. The *p*-values equal to or more than 0.05 indicate that the degree of the ncRNA synergistic competition networks obeys the power-law distribution. The Path and Density columns represent the –log10(*p*-value) of Student's *t*-test, and the *p*-values less than 0.05 suggest that the characteristic path lengths (or densities) of ncRNA synergistic competition networks are significantly shorter (or higher) than those of their corresponding random networks. (c) The similarity matrix showing the similarity between each pair of ncRNA synergistic competition networks across 31 malignant tumors. (d) The radar chart showing the percentages of ncRNA synergistic competition interactions predicted in different number of malignant tumors. (e) The bar chart showing the percentages of ncRNA synergistic competition interactions predicted in one malignant tumor.

Based on the topological features of the ncRNA synergistic competition networks, we further identify hub ncRNAs associated with malignant tumors. We have discovered that the number of hub ncRNAs vary across different malignant tumors (**Fig 4a**). In terms of the hub ncRNAs, the similarity between the 31 malignant tumors is [0.07 0.79] (**Fig S6** in **S1 File**). This result shows that the malignant tumors may have common hub ncRNAs. In addition, we have observed that the hub ncRNAs can significantly distinguish the high risk and the low risk groups of each malignant tumor (**Fig S7a** in **S1 File**), and the hub ncRNAs covering 96.77% malignant tumors are significantly enriched at least one of Disease, Cell markers and Cancer hallmarks terms (**Fig S7b** in **S1 File**, **S5 Data**). Different from the ncRNA synergistic competition networks, less than 13% hub ncRNAs are only in one malignant tumor but nearly 90% hub ncRNAs exist in at least two malignant tumors (**Fig 4b**). In particular, ~7.63% hub ncRNAs are even conserved in at least 16 malignant tumors. This result reveals that the hub ncRNAs of different malignant tumors tend to be conserved in their respective tumor microenvironments.

To understand whether there is a common core of hub ncRNAs across malignant tumors, we investigate the conserved hub ncRNAs occurred in at least 16 malignant tumors. Functional analysis indicates that the conserved hub ncRNAs have better performance than the baseline method in classifying the 31 malignant tumors (**Fig 4c**), and can significantly distinguish the high risk and the low risk groups of each malignant tumor (**Fig 4d**). This result shows that the conserved hub ncRNAs may act as potential diagnostic and prognostic biomarkers across malignant tumors. Moreover, the conserved hub ncRNAs are significantly enriched in diseases (**Fig 4e**), cell markers (**Fig 4f**) and cancer hallmarks (**Fig 4g**) common to several malignant tumors. This result suggests that the malignant tumors may have common cell markers and cancer hallmarks.

In addition to the conserved hub ncRNAs, the rewired hub ncRNAs switching from tumor to tumor are important to understand the heterogeneity of malignant tumors. We have found that the percentage of the rewired hub ncRNAs for each malignant tumor is less than 7%, revealing a small heterogeneity across malignant tumors in terms of hub ncRNAs (**Fig 4h**). Multi-class classification analysis shows that the rewired hub ncRNAs have better performance than the baseline method in classifying the 31 malignant tumors (**Fig S7c** in **S1 File**). Moreover, survival analysis displays that the rewired hub ncRNAs can significantly distinguish the high risk and the low risk groups of each malignant tumor (**Fig S7d** in **S1**

**File**). This result suggests that the rewired hub ncRNAs may also act as potential diagnostic and prognostic biomarkers across malignant tumors.




| Metrics | Conserved | Baseline |
|---|---|---|
| accuracy | 0.91 | 0.09 |
| average-precision | 0.95 | 0.17 |
| clp | 0.00 | 0.03 |
| coverage | 0.21 | 14.42 |
| F1 | 0.91 | 0.09 |
| hamming-loss | 0.01 | 0.06 |
| macro-AUC | 0.99 | 0.50 |
| macro-F1 | 0.89 | 0.01 |
| macro-precision | 0.89 | 0.00 |
| macro-recall | 0.90 | 0.03 |
| margin-loss | 0.21 | 14.42 |
| micro-AUC | 1.00 | 0.53 |
| micro-F1 | 0.91 | 0.09 |
| micro-precision | 0.90 | 0.09 |
| micro-recall | 0.92 | 0.09 |
| mlp | 0.00 | 0.97 |
| one-error | 0.09 | 0.91 |
| precision | 0.91 | 0.09 |
| ranking-loss | 0.01 | 0.48 |
| recall | 0.92 | 0.09 |
| subset-accuracy | 0.89 | 0.09 |
| wlp | 0.00 | 0.97 |

| Tumors | p-value | HR | HRlow95 | HRup95 |
|---|---|---|---|---|
| ACC | 5.33E-15 | 9.79 | 3.33 | 28.81 |
| BLCA | <2.22E-16 | 10.15 | 6.53 | 15.77 |
| BRCA | <2.22E-16 | 14.78 | 8.13 | 26.88 |
| CESC | <2.22E-16 | 8.80 | 3.99 | 19.42 |
| CHOL | 4.63E-09 | 9.31 | 2.08 | 41.74 |
| COAD | <2.22E-16 | 8.44 | 3.66 | 19.45 |
| DLBC | 6.51E-08 | 18.77 | 3.12 | 113.01 |
| ESCA | <2.22E-16 | 7.50 | 3.18 | 17.72 |
| HNSC | <2.22E-16 | 12.06 | 7.80 | 18.66 |
| KICH | 6.19E-07 | / | / | / |
| KIRC | <2.22E-16 | 15.26 | 7.05 | 33.00 |
| KIRP | <2.22E-16 | 11.18 | 4.04 | 30.94 |
| LGG | <2.22E-16 | 10.38 | 6.16 | 17.48 |
| LIHC | <2.22E-16 | 14.06 | 8.49 | 23.30 |
| LUAD | <2.22E-16 | 10.85 | 6.92 | 17.03 |
| LUSC | <2.22E-16 | 12.98 | 7.80 | 21.62 |
| MESO | <2.22E-16 | 7.31 | 2.99 | 17.90 |
| OV | <2.22E-16 | 7.49 | 4.75 | 11.81 |
| PAAD | <2.22E-16 | 7.69 | 3.83 | 15.44 |
| PCPG | 1.64E-05 | / | / | / |
| PRAD | 3.74E-10 | 23.38 | 5.30 | 103.27 |
| READ | 3.05E-10 | 10.00 | 2.38 | 42.11 |
| SARC | <2.22E-16 | 8.68 | 4.58 | 16.43 |
| SKCM | <2.22E-16 | 8.50 | 5.44 | 13.28 |
| STAD | <2.22E-16 | 10.95 | 6.16 | 19.47 |
| TGCT | 2.36E-03 | / | / | / |
| THCA | 7.27E-14 | / | / | / |
| THYM | 6.66E-09 | 20.61 | 3.23 | 131.69 |
| UCEC | <2.22E-16 | 64.84 | 35.96 | 116.91 |
| UCS | 1.11E-16 | 9.22 | 3.04 | 27.95 |
| UVM | 2.16E-14 | 11.05 | 3.19 | 38.26 |

**Fig 4. Hub analysis of ncRNA synergistic competition across malignant tumors.** (a) The number of hub ncRNAs in each malignant tumor. (b) The radar chart displaying the percentage of hub ncRNAs predicted in different number of malignant tumors. (c) Multi-class classification analysis of conserved hub ncRNAs in at least 16 malignant tumors. (d) Survival analysis of conserved hub ncRNAs in at least 16 malignant tumors. The symbol "/" stands for an infinite value. (e-g) Enrichment analysis (three categories

including disease, cell marker and cancer hallmark) of conserved hub ncRNAs in at least 16 malignant tumors. (h) The bar chart indicating the percentage of hub ncRNAs predicted in one malignant tumor.

**Synergistic competition ncRNAs are involved in drug resistance**

It is reported that ncRNAs play important roles in drug resistance, and the dysregulation of them can induce resistance to anti-cancer therapy and cause a failure of malignant tumor clinical treatment [53,54]. Drug enrichment analysis shows that the ncRNA synergistic competition networks of ~96.77% of the 31 malignant tumors and the hub ncRNAs of ~64.52% of the 31 malignant tumors are significantly enriched in at least one drug resistance term (**Fig 5a**, **S6 Data and S7 Data**). We have found that the ncRNA synergistic competition networks of 29 malignant tumors are involved in the drug *cisplatin* resistance (**Fig 5b**). The drug *cisplatin* is a well-known chemotherapy medication, and can be used to treat a number of cancers including bladder cancer, brain tumors, cervical cancer, esophageal cancer, head and neck cancer, lung cancer, mesothelioma, ovarian cancer, etc [55]. Moreover, the resistance of two drugs *L-685458* and *LBW242* is closely associated with the ncRNA synergistic competition networks of 4 malignant tumors (**Fig 5b**). The drug *L-685458* is a potent transition state analog (TSA) γ-secretase inhibitor (GSI) and is an attractive therapeutic candidate for Alzheimer Disease (AD) and cancers [56], and the drug *LBW242* is a potent and orally active proapoptotic IAP inhibitor and potentiates treat multiple myeloma and ovarian cancer [57,58]. Similar to the ncRNA synergistic competition networks, the top three enriched drugs of hub ncRNAs across malignant tumors are also *cisplatin*, *L-685458* and

*LBW242* (**Fig 5c**). This result suggests that the synergistic competition ncRNAs (i.e. ncRNAs involved in the synergistic competition) across malignant tumors are involved in drug resistance, and may act as targets for anti-cancer therapy.

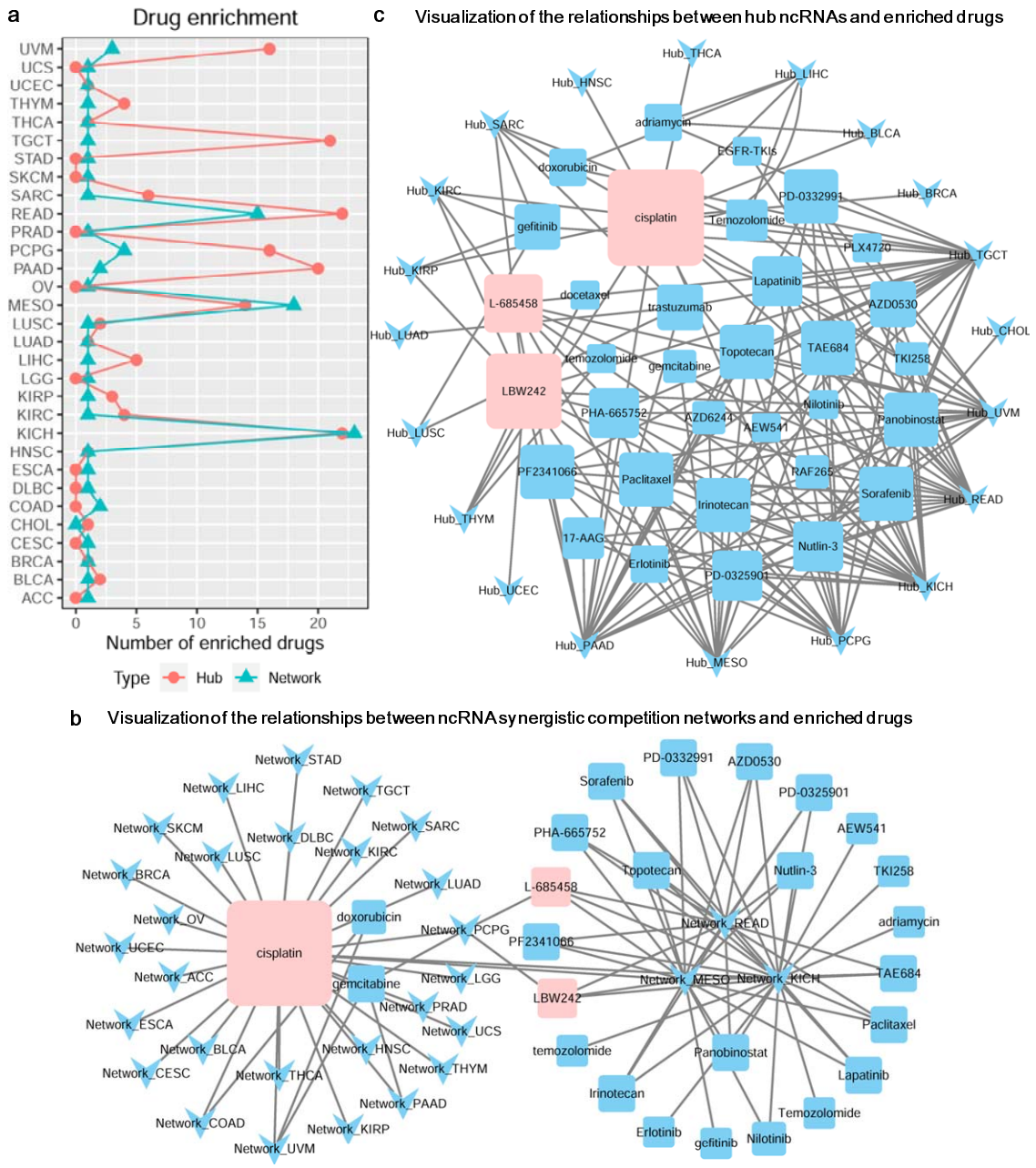


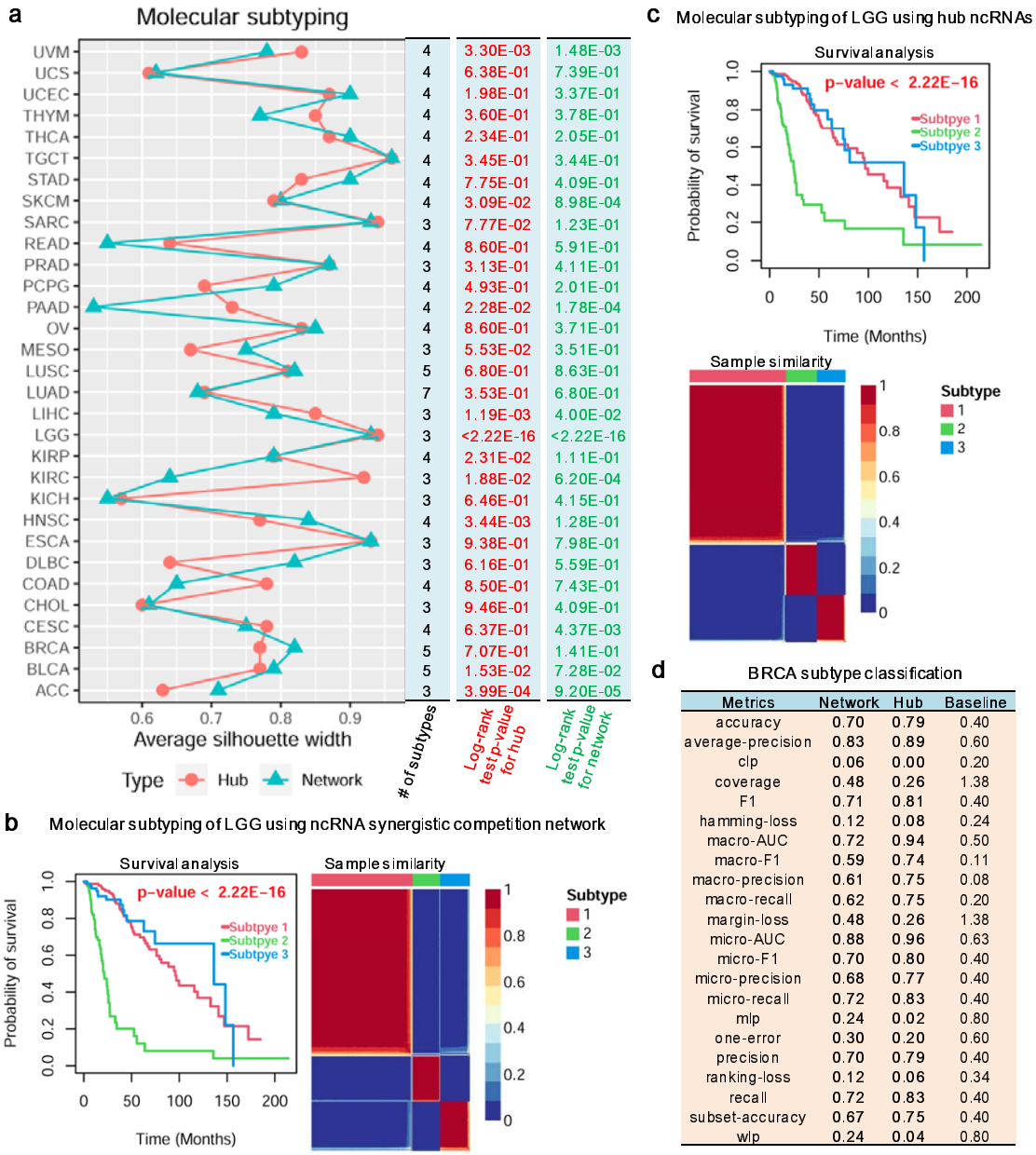

**Fig 5. Drug resistance analysis of ncRNA synergistic competition networks and hub ncRNAs across malignant tumors.** (a) The number of significantly enriched drugs for each malignant tumor. (b) Relationships between ncRNA synergistic competition networks and enriched drugs. For example, the

node “Network_ACC” denotes the ncRNA synergistic competition network in ACC. Larger drug nodes denote more links with ncRNA synergistic competition networks in malignant cancers. (c) Relationships between hub ncRNAs and enriched drugs. For example, the node “Hub_ACC” denotes the hub ncRNAs in ACC. Larger drug nodes denote more links with hub ncRNAs in malignant cancers.

**Synergistic competition ncRNAs contribute to distinguishing molecular subtypes of malignant tumors**

Malignant tumors are a collection of heterogeneous diseases with both phenotypic and genetic heterogeneity, and a malignant tumor usually possess multiple distinct molecular subtypes [59]. Molecular subtyping analysis of the synergistic competition ncRNAs (i.e. ncRNAs involved in the synergistic competition) shows that the average silhouette width of the molecular subtyping of ~90.32% of the 31 malignant tumors is equal to or greater than 0.60 (**Fig 6a**), indicating that the synergistic competition ncRNAs can well assign malignant tumor samples into distinct molecular subtypes. In the case of using hub ncRNAs, the molecular subtypes of ~32.26% of the 31 malignant tumors can be significantly distinguished (Log-rank test $p$-value < 0.05, **Fig 6a**). Moreover, in the case of using ncRNA synergistic competition networks, the molecular subtypes of ~25.81% of the 31 malignant tumors can be significantly distinguished (Log-rank test $p$-value < 0.05, **Fig 6a**). For seven malignant tumor types, including Adrenocortical carcinoma (ACC), Kidney renal clear cell carcinoma (KIRC), Brain Lower Grade Glioma (LGG), Liver hepatocellular carcinoma (LIHC), Pancreatic adenocarcinoma (PAAD), Skin Cutaneous Melanoma (SKCM) and Uveal Melanoma (UVM), the hub ncRNAs and ncRNA synergistic competition networks all have contributions to

distinguish molecular subtypes of them. For example, in the case of using hub ncRNAs and ncRNA synergistic competition networks, LGG can be divided into three distinct molecular subtypes (**Fig 6b** and **Fig 6c**). This result reveals that the synergistic competition ncRNAs contribute to distinguishing molecular subtypes of malignant tumors, and may act as prognostic biomarkers in several malignant tumors. Since the molecular subtypes (including Luminal A, Luminal B, HER-2, Basal, Normal-like) of Breast invasive carcinoma (BRCA) are mostly accepted by researchers, we focus on studying the BRCA subtype classification performance by using hub ncRNAs and ncRNA synergistic competition networks in BRCA. Multi-class classification analysis shows that the synergistic competition ncRNAs have better performance than the baseline method in classifying BRCA subtypes (**Fig 6d**), suggesting that the synergistic competition ncRNAs may act as potential diagnostic biomarkers of BRCA.




d

BRCA subtype classification

| Metrics | Network | Hub | Baseline |
|---|---|---|---|
| accuracy | 0.70 | 0.79 | 0.40 |
| average-precision | 0.83 | 0.89 | 0.60 |
| clp | 0.06 | 0.00 | 0.20 |
| coverage | 0.48 | 0.26 | 1.38 |
| F1 | 0.71 | 0.81 | 0.40 |
| hamming-loss | 0.12 | 0.08 | 0.24 |
| macro-AUC | 0.72 | 0.94 | 0.50 |
| macro-F1 | 0.59 | 0.74 | 0.11 |
| macro-precision | 0.61 | 0.75 | 0.08 |
| macro-recall | 0.62 | 0.75 | 0.20 |
| margin-loss | 0.48 | 0.26 | 1.38 |
| micro-AUC | 0.88 | 0.96 | 0.63 |
| micro-F1 | 0.70 | 0.80 | 0.40 |
| micro-precision | 0.68 | 0.77 | 0.40 |
| micro-recall | 0.72 | 0.83 | 0.40 |
| mlp | 0.24 | 0.02 | 0.80 |
| one-error | 0.30 | 0.20 | 0.60 |
| precision | 0.70 | 0.79 | 0.40 |
| ranking-loss | 0.12 | 0.06 | 0.34 |
| recall | 0.72 | 0.83 | 0.40 |
| subset-accuracy | 0.67 | 0.75 | 0.40 |
| wlp | 0.24 | 0.04 | 0.80 |

**Fig 6. Molecular subtype analysis of malignant tumors.** (a) Molecular subtyping of the 31 malignant tumors using ncRNA synergistic competition networks and hub ncRNAs. Higher average silhouette width indicates that a malignant tumor sample is better matched with the identified molecular subtypes. (b) Molecular subtyping of LGG using ncRNA synergistic competition network. (c) Molecular subtyping of LGG using hub ncRNAs. (d) Classification analysis of BRCA subtypes.

**Synergistic competition ncRNAs participate in immune regulation**

As regulatory elements in immune system, ncRNAs can influence immune development and function, leading to immune diseases including autoimmunity and haematological cancers [60,61]. Immune regulation analysis reveals that the ncRNA synergistic competition networks of all the 31 malignant tumors and the hub ncRNAs of 30 malignant tumors are significantly involved in at least one immune-related pathway (**Fig 7a**, **S8 Data and S9 Data**). For instance, the ncRNA synergistic competition networks of the 31 malignant tumors are closely associated with six immune-related pathways (*Antigen Processing and Presentation*, *Cytokine Receptors*, *Antimicrobials*, *Cytokines*, *Natural Killer Cell Cytotoxicity*, *TCR signaling Pathway*) (**Fig 7c**). In addition, the hub ncRNAs of 29 malignant tumors are closely related to two immune-related pathways (*Antigen Processing and Presentation*, *Cytokine Receptors*) (**Fig 7d**). Moreover, the ncRNA synergistic competition networks and the hub ncRNAs of 31 malignant tumors are significantly associated with at least 40 immune cells (**Fig 7b**, **S10 Data and S11 Data**). This result suggests that the synergistic competition ncRNAs (i.e. ncRNAs involved in the synergistic competition) take part in immune regulation, and the dysregulation of them may influence the occurrence and development of immune diseases including malignant tumors.

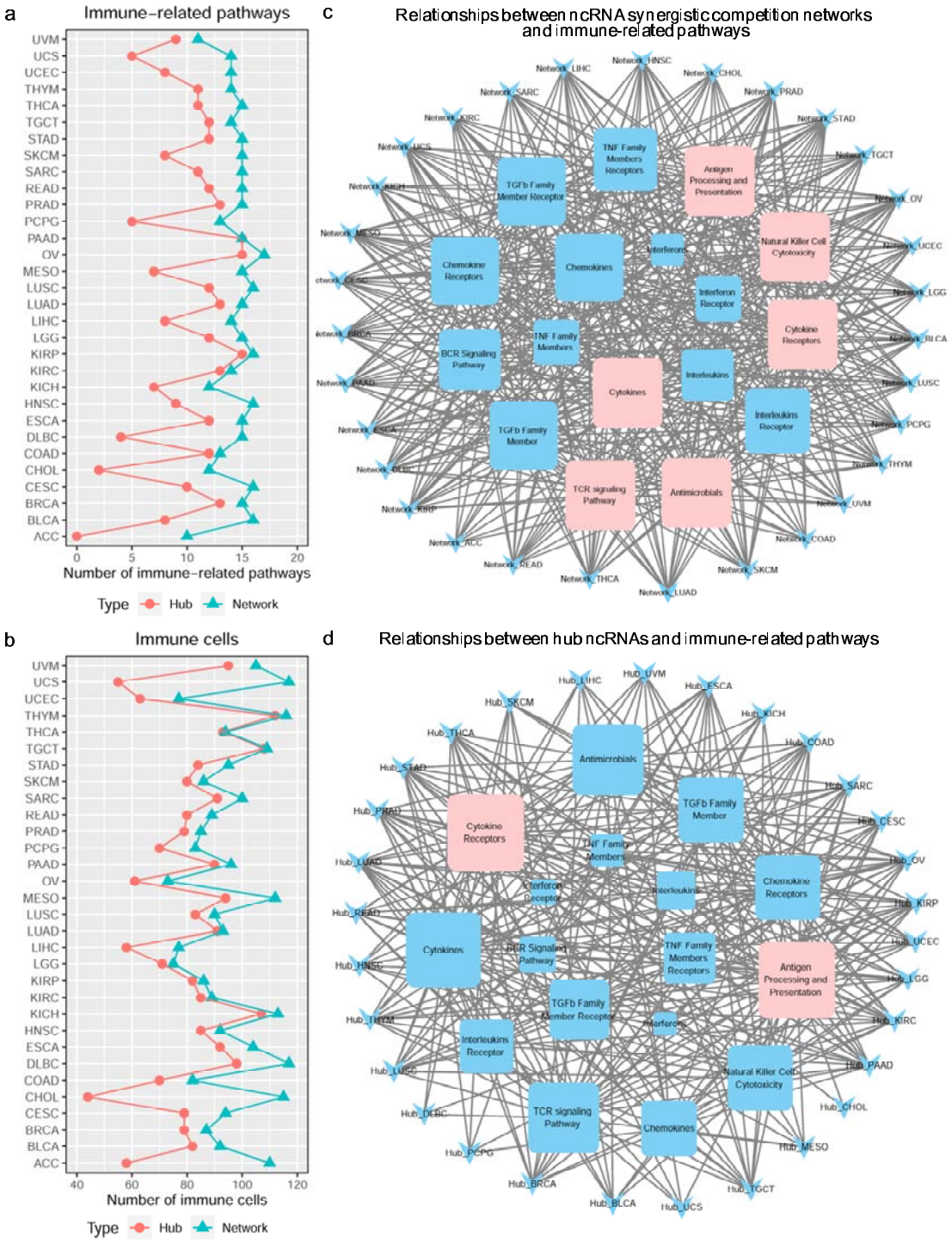



**Fig 7. Immune regulation analysis of synergistic competition ncRNAs across malignant tumors.** (a) The number of significant immune-related pathways associated with each malignant tumor. (b) The number of significant immune cells associated with each malignant tumor. (c) Relationships between ncRNA synergistic competition networks and immune-related pathways. For example, the node "Network_ACC" denotes the ncRNA synergistic competition network in ACC. Larger immune-related pathway nodes denote more links with ncRNA synergistic competition networks in malignant cancers. (d)

Relationships between hub ncRNAs and immune-related pathways. For example, the node "Hub_ACC" denotes the hub ncRNAs in ACC. Larger immune-related pathway nodes denote more links with hub ncRNAs in malignant cancers.

**SCOMdb: a web-based resource for ncRNA synergistic competition in Pan-cancer**

To provide a convenient and web-accessible resource of ncRNA synergistic competition landscape across malignant tumors for biologists, we have developed a comprehensive and interactive database, SCOMdb (www.comblab.cn/SCOMdb/). By using this web-based resource, users can browse the ceRNA interactions in specific malignant tumor of interest (**Fig 8a**). It also provides the ncRNA synergistic competition interactions in malignant tumors (**Fig 8b**). In addition, users can query the hub ncRNAs of interest across malignant tumors (**Fig 8c**). All the data in this work can be downloaded for downstream analysis (**Fig 8d**). To help users apply SCOM to new datasets, we have also provided the SCOM R package for inferring ncRNA synergistic competition by integrating sample-matched gene (ncRNA and mRNA) expression data and putative miRNA-target interactions (**Fig 8e**). The provided resources will be continuously updated, and should serve as a valuable resource for biologists interested in investigating ncRNA synergistic competition across malignant tumors.

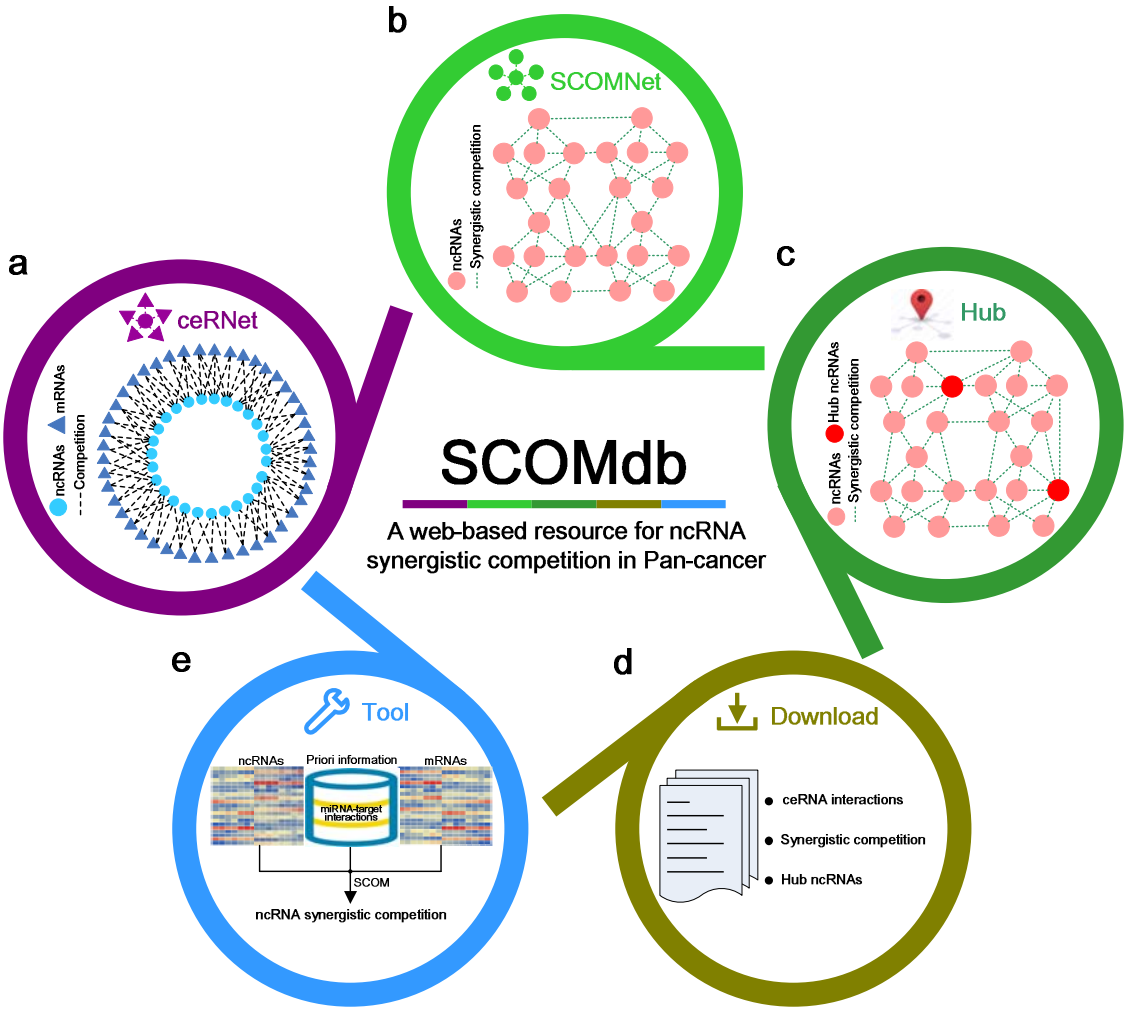



**Fig 8. Diagram of SCOMdb database.** (a) The search of ceRNA regulation in Pan-cancer. (b) The search of ncRNA synergistic competition in Pan-cancer. (c) The search of hub ncRNAs in Pan-cancer. (d) All resources are provided to be downloaded for further analysis. (e) The SCOM method is available for users to study ncRNA synergistic competition by using their own datasets.

## Conclusions and discussion

Abundant evidence reveals that ncRNAs can serve as important regulators across malignant tumors. However, from the perspective of synergistic competition, ncRNA regulation is still unearthed. In this work, we present the *ncRNA synergistic competition hypothesis* and the SCOM framework, and further elaborate the application of the SCOM framework to

comprehensively infer the synergistic competition ncRNAs that potentially act as carcinogenic biomarkers. From the perspective of synergistic competition, we have shown that the synergistic competition ncRNAs are potential diagnostic and prognostic biomarkers of malignant tumors. Moreover, the synergistic competition ncRNAs are likely to involve in drug resistance, contribute to distinguishing molecular subtypes of malignant tumors, and participate in immune regulation. The web-based database SCOMdb (www.comblab.cn/SCOMdb/) provides a valuable resource for exploring ncRNA regulation across malignant tumors.

As important regulators of gene expression, ncRNAs play important roles in the physiology and development of malignant tumors. Yet, the experimentally validated carcinogenic biomarkers are far from complete. Thus, it is challenging to confirm the accuracy of the findings by the SCOM framework. To show the effectiveness of SCOM, we conduct a series of functional analyses including enrichment analysis, molecular subtyping analysis, multi-class classification analysis, survival analysis, and immune regulation analysis for the synergistic competition ncRNAs. We have found that the conserved ncRNA synergistic competition ncRNAs and hub ncRNAs are significantly enriched in diseases, cell markers and cancer hallmarks. In addition, the conserved and rewired hub ncRNAs may act as potential diagnostic and prognostic biomarkers across malignant tumors. Moreover, the synergistic competition ncRNAs are likely to be involved in drug resistance, contribute to

distinguishing molecular subtypes of malignant tumors, and participate in immune regulation. These results indicate that the SCOM framework can effectively infer the synergistic competition ncRNAs with biological significance. Although experimentally validated carcinogenic biomarkers are limited, we have discovered that a significant number of synergistic competition ncRNAs are confirmed as carcinogenic biomarkers by the literature (**Fig S8** in **S1 File**, **S12 Data**).

The findings in this paper have provided several biological insights although the identified synergistic competition ncRNAs need to be further confirmed by subsequent biological experiments. Firstly, not all the ncRNA synergistic competition networks follow a power law distribution. For example, the ncRNA synergistic competition networks in Kidney Chromophobe (KICH), Lung adenocarcinoma (LUAD), Lung squamous cell carcinoma (LUSC) and Rectum adenocarcinoma (READ) do not obey the power-law distribution. This result indicates that the synergistic competition ncRNAs in these four malignant tumors might equally interact with each other. Secondly, a majority of the ncRNA interactions involved in the synergistic competition are rewired, but a minority of hub ncRNAs is rewired. This result suggests that malignant tumors are likely to share hub ncRNAs rather than the ncRNA interactions involved in the synergistic competition. Thirdly, the similarity of hub ncRNAs between the 31 malignant tumors is significantly larger than that of ncRNA synergistic competition networks ($p$-value < 2.20E-16 with paired Student's $t$-test).

Moreover, the percentage of rewired ncRNA synergistic competition interactions is significantly larger than that of rewired hub ncRNAs (*p*-value < 2.20E-16 with paired Student's *t*-test). This result reveals that although malignant tumors have great differences in ncRNA synergistic competition networks, the hub ncRNAs between malignant tumors tend to be similar. Fourthly, the synergistic competition ncRNAs can distinguish five BRCA subtypes with accuracy equal to or more than 70%. This result shows that the synergistic competition ncRNAs might be good diagnostic biomarkers in BRCA. Fifthly, we have used in this study the transcriptomics data in TCGA [27], which is one of the most well-known cancer genomics programs. However, the SCOM framework can also be applied to the transcriptomics data in other sources (e.g. CCLE [62] and MET500 [63]) to investigate ncRNA synergistic competition. Finally, SCOM is not limited to the exploration of lncRNA and pseudogene synergistic competition. It can also be used to infer the synergistic competition of other ncRNAs (e.g. circRNAs) that are important modulators to malignant tumors.

Prediction of ncRNA-related ceRNA network is a critical step to infer ncRNA synergistic competition. Yet, there is still a lack of study on the prediction of ncRNA-related ceRNA networks. In the current work, we apply three criteria (significant sharing of miRNAs, significant positive correlation, and significant sensitive correlation conditioning on shared miRNAs) to predict ncRNA-related ceRNA networks in 31 malignant tumors. Since the three

criteria have been successfully used to investigate ceRNA regulation [33,35,64–66], the ncRNA-related ceRNA network predictions are promising. However, evaluating whether a ncRNA can act as a ceRNA to compete with mRNAs is still challenging. For example, the stoichiometry of miRNAs and ncRNAs is considered as a vital factor to predict ceRNA interactions of a ncRNA of interest [67]. With deeper understanding of ceRNA mechanisms, we can more accurately predict ncRNA-related ceRNA networks to improve the inference of ncRNA synergistic competition and accelerate the discovery of carcinogenic biomarkers.

Given gene expression data and miRNA-target interactions, SCOM constructs the landscape of ncRNA synergistic competition across malignant tumors at the bulk level. Recently, increasingly advanced single-cell and spatial RNA sequencing technology has opened a way for exploring ncRNA synergistic competition at the single-cell level. With the advance of single-cell and spatial RNA sequencing technology, SCOM will contribute to uncover more carcinogenic biomarkers.

ncRNAs are emerging as potential carcinogenic biomarkers. It is important for us to understand the regulation of ncRNAs in malignant tumors. In-depth study of synergistic competition ncRNAs inferred in this study would contribute to the design of reliable diagnosis and treatment biomarkers for human malignant tumors.

## References

1. Hanahan D. Hallmarks of cancer: new dimensions. Cancer Discov. 2022;12: 31–46. doi:10.1158/2159-8290.CD-21-1059

2. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2021;71: 209–249. doi:10.3322/caac.21660
3. Weinberg RA. The biology of cancer, 2nd Edition. Second Edition. New York;
4. Sparmann A, van Lohuizen M. Polycomb silencers control cell fate, development and cancer. Nat Rev Cancer. 2006;6: 846–856. doi:10.1038/nrc1991
5. Dahariya S, Paddibhatla I, Kumar S, Raghuwanshi S, Pallepati A, Gutti RK. Long non-coding RNA: Classification, biogenesis and functions in blood cells. Mol Immunol. 2019;112: 82–92. doi:10.1016/j.molimm.2019.04.011
6. Esteller M. Non-coding RNAs in human disease. Nat Rev Genet. 2011;12: 861–874. doi:10.1038/nrg3074
7. Anastasiadou E, Jacob LS, Slack FJ. Non-coding RNA networks in cancer. Nat Rev Cancer. 2018;18: 5–18. doi:10.1038/nrc.2017.99
8. Fu XD. Non-coding RNA: a new frontier in regulatory biology. Natl Sci Rev. 2014;1: 190–204. doi:10.1093/nsr/nwu008
9. Salmena L, Poliseno L, Tay Y, Kats L, Pandolfi PP. A ceRNA hypothesis: the Rosetta Stone of a hidden RNA language? Cell. 2011;146: 353–358. doi:10.1016/j.cell.2011.07.014
10. Smillie CL, Sirey T, Ponting CP. Complexities of post-transcriptional regulation and the modeling of ceRNA crosstalk. Crit Rev Biochem Mol Biol. 2018;53: 231–245. doi:10.1080/10409238.2018.1447542
11. Tay Y, Rinn J, Pandolfi PP. The multilayered complexity of ceRNA crosstalk and competition. Nature. 2014;505: 344–352. doi:10.1038/nature12986
12. Wang Y, Hou J, He D, Sun M, Zhang P, Yu Y, et al. The emerging function and mechanism of ceRNAs in cancer. Trends Genet. 2016;32: 211–224. doi:10.1016/j.tig.2016.02.001
13. Qi X, Zhang D-H, Wu N, Xiao J-H, Wang X, Ma W. ceRNA in cancer: possible functions and clinical implications. J Med Genet. 2015;52: 710–718. doi:10.1136/jmedgenet-2015-103334
14. Peng F, Li T-T, Wang K-L, Xiao G-Q, Wang J-H, Zhao H-D, et al. H19/let-7/LIN28 reciprocal negative regulatory circuit promotes breast cancer stem cell maintenance. Cell Death Dis. 2017;8: e2569. doi:10.1038/cddis.2016.438
15. Cheng Z, Yu C, Cui S, Wang H, Jin H, Wang C, et al. circTP63 functions as a ceRNA to promote lung squamous cell carcinoma progression by upregulating FOXM1. Nat Commun. 2019;10: 3200. doi:10.1038/s41467-019-11162-4

16. Poliseno L, Salmena L, Zhang J, Carver B, Haveman WJ, Pandolfi PP. A coding-independent function of gene and pseudogene mRNAs regulates tumour biology. Nature. 2010;465: 1033–1038. doi:10.1038/nature09144
17. Arun K, Arunkumar G, Bennet D, Chandramohan SM, Murugan AK, Munirajan AK. Comprehensive analysis of aberrantly expressed lncRNAs and construction of ceRNA network in gastric cancer. Oncotarget. 2018;9: 18386–18399. doi:10.18632/oncotarget.24841
18. Xiong JM, Liu H, Chen J, Zou QQ, Wang YYJ, Bi G. Curcumin nicotinate suppresses abdominal aortic aneurysm pyroptosis via lncRNA PVT1/miR-26a/KLF4 axis through regulating the PI3K/AKT signaling pathway. Toxicol Res (Camb). 2021;10: 651–661. doi:10.1093/toxres/tfab041
19. Huang M, Zhong Z, Lv M, Shu J, Tian Q, Chen J. Comprehensive analysis of differentially expressed profiles of lncRNAs and circRNAs with associated co-expression and ceRNA networks in bladder carcinoma. Oncotarget. 2016;7: 47186–47200. doi:10.18632/oncotarget.9706
20. Sun X, Luo L, Li J. LncRNA MALAT1 facilitates BM-MSCs differentiation into endothelial cells via targeting miR-206/VEGFA axis. Cell Cycle. 2020;19: 3018–3028. doi:10.1080/15384101.2020.1829799
21. Lai SW, Chen MY, Bamodu OA, Hsieh M-S, Huang T-Y, Yeh C-T, et al. Exosomal lncRNA PVT1/VEGFA axis promotes colon cancer metastasis and stemness by downregulation of tumor suppressor miR-152-3p. Oxid Med Cell Longev. 2021;2021: 9959807. doi:10.1155/2021/9959807
22. Chen L, Yao H, Wang K, Liu X. Long non-coding RNA MALAT1 regulates ZEB1 expression by sponging miR-143-3p and promotes hepatocellular carcinoma progression. J Cell Biochem. 2017;118: 4836–4843. doi:10.1002/jcb.26158
23. Yang T, Zhou H, Liu P, Yan L, Yao W, Chen K, et al. lncRNA PVT1 and its splicing variant function as competing endogenous RNA to regulate clear cell renal cell carcinoma progression. Oncotarget. 2017;8: 85353–85367. doi:10.18632/oncotarget.19743
24. Wang Y, Zhou Y, Yang Z, Chen B, Huang W, Liu Y, et al. MiR-204/ZEB2 axis functions as key mediator for MALAT1-induced epithelial-mesenchymal transition in breast cancer. Tumour Biol. 2017;39: 1010428317690998. doi:10.1177/1010428317690998
25. Chen SW, Zhu J, Ma J, Zhang J, Zuo S, Chen G-W, et al. Overexpression of long non-coding RNA H19 is associated with unfavorable prognosis in patients with colorectal cancer and increased proliferation and migration in colon cancer cells. Oncol Lett. 2017;14: 2446–2452. doi:10.3892/ol.2017.6390

26. Choobdar S, Ahsen ME, Crawford J, Tomasoni M, Fang T, Lamparter D, et al. Assessment of network module identification across complex diseases. Nat Methods. 2019;16: 843–852. doi:10.1038/s41592-019-0509-5
27. Weinstein JN, Collisson EA, Mills GB, Shaw KRM, Ozenberger BA, Ellrott K, et al. The Cancer Genome Atlas Pan-Cancer analysis project. Nat Genet. 2013;45: 1113–1120. doi:10.1038/ng.2764
28. Huang H, Lin Y-C, Cui S, Huang Y, Tang Y, Xu J, et al. miRTarBase update 2022: an informative resource for experimentally validated miRNA-target interactions. Nucleic Acids Res. 2022;50: D222–D230. doi:10.1093/nar/gkab1079
29. Karagkouni D, Paraskevopoulou MD, Chatzopoulos S, Vlachos IS, Tastsoglou S, Kanellos I, et al. DIANA-TarBase v8: a decade-long collection of experimentally supported miRNA-gene interactions. Nucleic Acids Res. 2018;46: D239–D245. doi:10.1093/nar/gkx1141
30. Paraskevopoulou MD, Vlachos IS, Karagkouni D, Georgakilas G, Kanellos I, Vergoulis T, et al. DIANA-LncBase v2: indexing microRNA targets on non-coding transcripts. Nucleic Acids Res. 2016;44: D231-238. doi:10.1093/nar/gkv1270
31. Teng X, Chen X, Xue H, Tang Y, Zhang P, Kang Q, et al. NPInter v4.0: an integrated database of ncRNA interactions. Nucleic Acids Res. 2020;48: D160–D165. doi:10.1093/nar/gkz969
32. Li JH, Liu S, Zhou H, Qu LH, Yang JH. starBase v2.0: decoding miRNA-ceRNA, miRNA-ncRNA and protein-RNA interaction networks from large-scale CLIP-Seq data. Nucleic Acids Res. 2014;42: D92-97. doi:10.1093/nar/gkt1248
33. List M, Dehghani Amirabad A, Kostka D, Schulz MH. Large-scale inference of competing endogenous RNA networks with sparse partial correlation. Bioinformatics. 2019;35: i596–i604. doi:10.1093/bioinformatics/btz314
34. Pearson K. Notes on the history of correlation. Biometrika. 1920;13: 25–45. doi:10.2307/2331722
35. Paci P, Colombo T, Farina L. Computational analysis identifies a sponge interaction network between long non-coding RNAs and messenger RNAs in human breast cancer. BMC Syst Biol. 2014;8: 83. doi:10.1186/1752-0509-8-83
36. Telesford QK, Joyce KE, Hayasaka S, Burdette JH, Laurienti PJ. The ubiquity of small-world networks. Brain Connect. 2011;1: 367–375. doi:10.1089/brain.2011.0038
37. Zaidi F. Small world networks and clustered small world networks with random connectivity. Social Network Analysis and Mining. 2013;3: 51–63. doi:10.1007/s13278-012-0052-1
38. Csardi G, Nepusz T. The igraph software package for complex network research. InterJournal, Complex Systems. 2006;1695: 1–9.

39. Massey FJ. The Kolmogorov-Smirnov test for goodness of fit. Journal of the American Statistical Association. 1951;46: 68–78. doi:10.2307/2280095
40. Student. The probable error of a mean. Biometrika. 1908;6: 1–25. doi:10.2307/2331554
41. Tsoumakas G, Katakis I, Vlahavas I. Mining multi-label data. Data mining and knowledge discovery handbook Springer, Boston, MA. Springer, Boston, MA; 2009. pp. 667–685.
42. Rivolli A, Carvalho ACPLF de. The utiml package: multi-label classification in R. The R Journal. 2018;10: 24–37. doi:10.32614/RJ-2018-041
43. Chang CC, Lin CJ. LIBSVM: a library for support vector machines. ACM Transactions on Intelligent Systems and Technology. 2011;2: 1–27.
44. Meyer D, Dimitriadou E, Hornik K, Weingessel A, Leisch F, Chang C-C, et al. e1071: misc functions of the department of statistics, probability theory group (Formerly: E1071), TU Wien. Available: https://CRAN.R-project.org/package=e1071
45. Metz J, de Abreu LF, Cherman EA, Monard MC. On the estimation of predictive evaluation measure baselines for multi-label learning. Ibero-American Conference on Artificial Intelligence, Cartagena de Indias, Colombia. Springer; 2012. pp. 189–198.
46. Andersen PK, Gill RD. Cox's regression model for counting processes: a large sample study. Ann Statist. 1982;10: 1100–1120. doi:10.1214/aos/1176345976
47. Terry M. Therneau, Patricia M. Grambsch. Modeling survival data: extending the Cox model. New York: Springer; 2000.
48. Chen J, Zhang J, Gao Y, Li Y, Feng C, Song C, et al. LncSEA: a platform for long non-coding RNA related sets and enrichment analysis. Nucleic Acids Res. 2021;49: D969–D980. doi:10.1093/nar/gkaa806
49. Xu T, Le TD, Liu L, Su N, Wang R, Sun B, et al. CancerSubtypes: an R/Bioconductor package for molecular cancer subtype identification, validation and visualization. Bioinformatics. 2017;33: 3131–3133. doi:10.1093/bioinformatics/btx378
50. Wang B, Mezlini AM, Demir F, Fiume M, Tu Z, Brudno M, et al. Similarity network fusion for aggregating data types on a genomic scale. Nat Methods. 2014;11: 333–337. doi:10.1038/nmeth.2810
51. Monti S, Tamayo P, Mesirov J, Golub T. Consensus clustering: a resampling-based method for class discovery and visualization of gene expression microarray data. Machine Learning. 2003;52: 91–118. doi:10.1023/A:1023949509487
52. Jiang T, Zhou W, Chang Z, Zou H, Bai J, Sun Q, et al. ImmReg: the regulon atlas of immune-related pathways across cancer types. Nucleic Acids Res. 2021;49: 12106–12118. doi:10.1093/nar/gkab1041
53. Si W, Shen J, Zheng H, Fan W. The role and mechanisms of action of microRNAs in cancer drug resistance. Clin Epigenetics. 2019;11: 25. doi:10.1186/s13148-018-0587-8

54. Smallegan MJ, Rinn JL. Linking long noncoding RNA to drug resistance. Proc Natl Acad Sci U S A. 2019;116: 21963–21965. doi:10.1073/pnas.1915690116
55. Wishart DS, Feunang YD, Guo AC, Lo EJ, Marcu A, Grant JR, et al. DrugBank 5.0: a major update to the DrugBank database for 2018. Nucleic Acids Res. 2018;46: D1074–D1082. doi:10.1093/nar/gkx1037
56. Yang G, Zhou R, Guo X, Yan C, Lei J, Shi Y. Structural basis of γ-secretase inhibition and modulation by small molecule drugs. Cell. 2021;184: 521-533.e14. doi:10.1016/j.cell.2020.11.049
57. Weisberg E, Kung AL, Wright RD, Moreno D, Catley L, Ray A, et al. Potentiation of antileukemic therapies by Smac mimetic, LBW242: effects on mutant FLT3-expressing cells. Mol Cancer Ther. 2007;6: 1951–1961. doi:10.1158/1535-7163.MCT-06-0810
58. Petrucci E, Pasquini L, Bernabei M, Saulle E, Biffoni M, Accarpio F, et al. A small molecule SMAC mimic LBW242 potentiates TRAIL- and anticancer drug-mediated cell death of ovarian cancer cells. PLoS One. 2012;7: e35073. doi:10.1371/journal.pone.0035073
59. Zhao L, Lee VHF, Ng MK, Yan H, Bijlsma MF. Molecular subtyping of cancer: current status and moving toward clinical applications. Brief Bioinform. 2019;20: 572–584. doi:10.1093/bib/bby026
60. Mehta A, Baltimore D. MicroRNAs as regulatory elements in immune system logic. Nat Rev Immunol. 2016;16: 279–294. doi:10.1038/nri.2016.40
61. Peltier DC, Roberts A, Reddy P. LNCing RNA to immunity. Trends Immunol. 2022;43: 478–495. doi:10.1016/j.it.2022.04.002
62. Barretina J, Caponigro G, Stransky N, Venkatesan K, Margolin AA, Kim S, et al. The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. Nature. 2012;483: 603–607. doi:10.1038/nature11003
63. Robinson DR, Wu Y-M, Lonigro RJ, Vats P, Cobain E, Everett J, et al. Integrative clinical genomics of metastatic cancer. Nature. 2017;548: 297–303. doi:10.1038/nature23306
64. Le TD, Zhang J, Liu L, Li J. Computational methods for identifying miRNA sponge interactions. Brief Bioinform. 2017;18: 577–590. doi:10.1093/bib/bbw042
65. Li Y, Jin X, Wang Z, Li L, Chen H, Lin X, et al. Systematic review of computational methods for identifying miRNA-mediated RNA-RNA crosstalk. Brief Bioinform. 2019;20: 1193–1204. doi:10.1093/bib/bbx137
66. Zhang J, Liu L, Zhang W, Li X, Zhao C, Li S, et al. miRspongeR 2.0: an enhanced R package for exploring miRNA sponge regulation. Bioinformatics Advances. 2022;2: vbac063. doi:10.1093/bioadv/vbac063
67. Thomson DW, Dinger ME. Endogenous microRNA sponges: evidence and controversy. Nat Rev Genet. 2016;17: 272–283. doi:10.1038/nrg.2016.20

## Supporting information

**S1 Data. The matched tumor types in Pan-cancer dataset.**

**S2 Data. Clinical information of the matched 8001 tumor samples.**

**S3 Data. The molecular subtypes of malignant tumors in Pan-cancer dataset.**

**S4 Data. Enrichment analysis of the conserved ncRNA synergistic competition in at least two malignant tumors.**

**S5 Data. Enrichment analysis of hub ncRNAs across malignant tumors.**

**S6 Data. Drug enrichment of ncRNA synergistic competition networks.**

**S7 Data. Drug enrichment of hub ncRNAs.**

**S8 Data. Relationships between ncRNA synergistic competition networks and immune-related pathways across malignant tumors.**

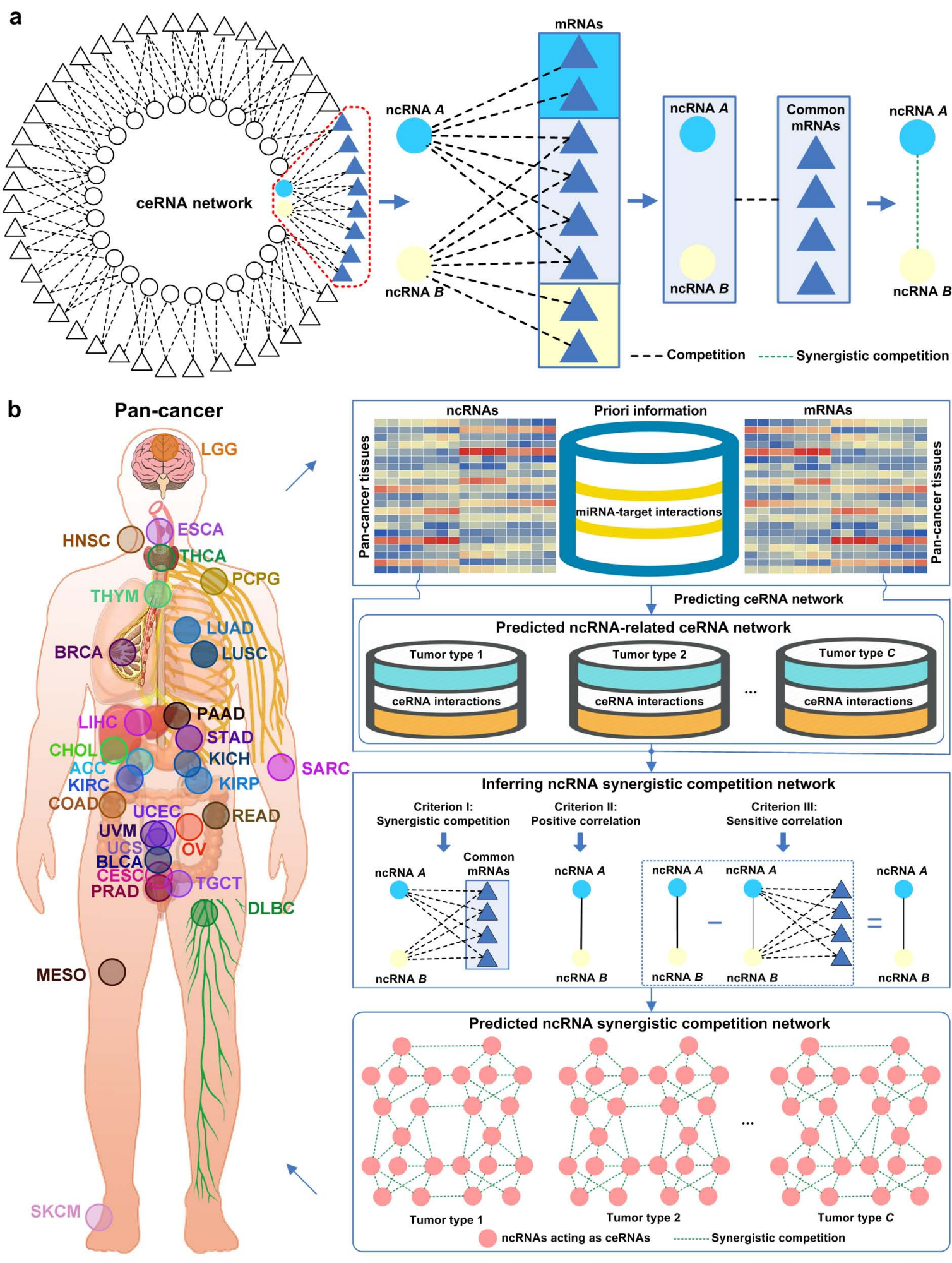
**S9 Data. Relationships between hub ncRNAs and immune-related pathways across malignant tumors.**

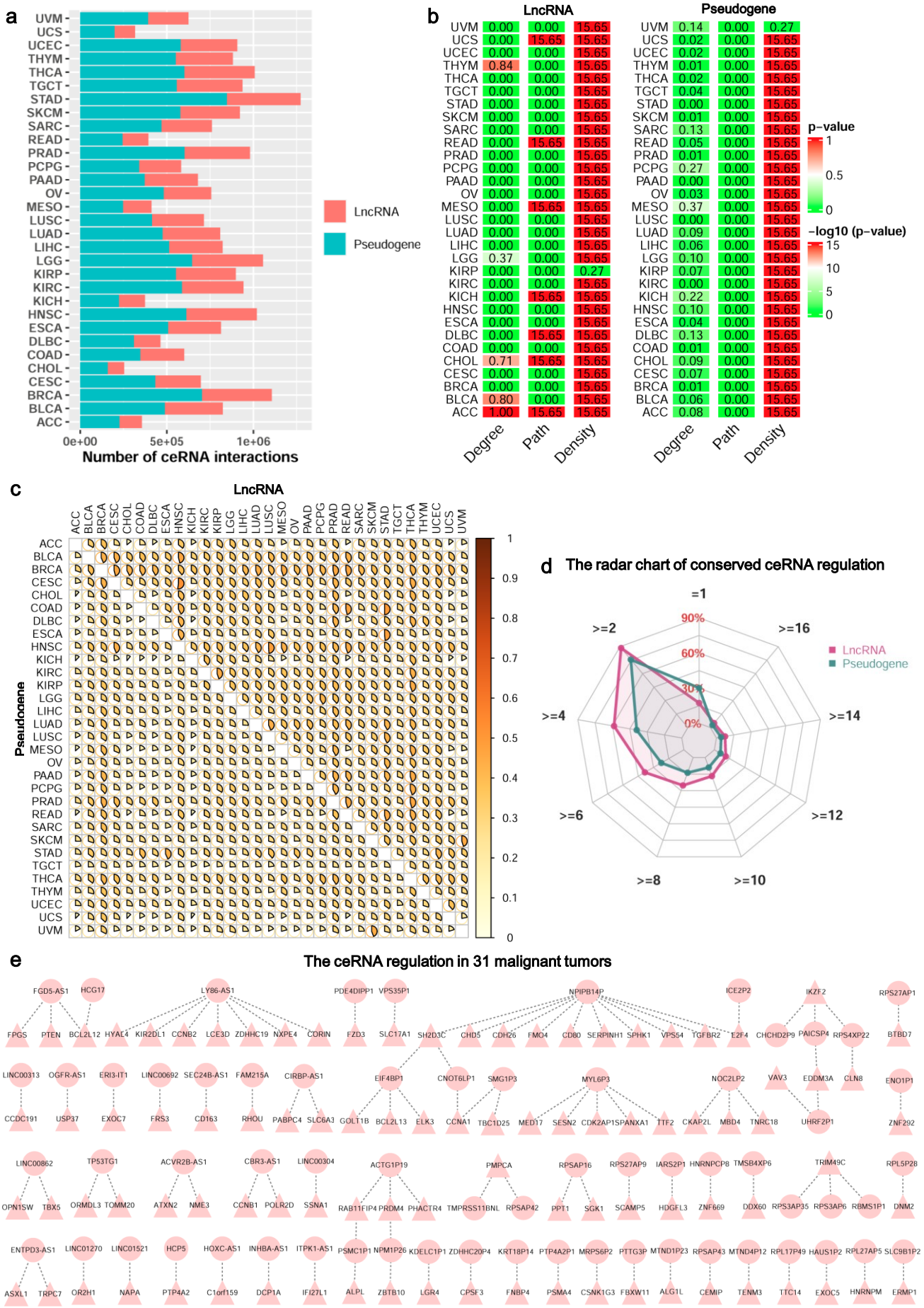
**S10 Data. Relationships between ncRNA synergistic competition networks and immune cells across malignant tumors.**

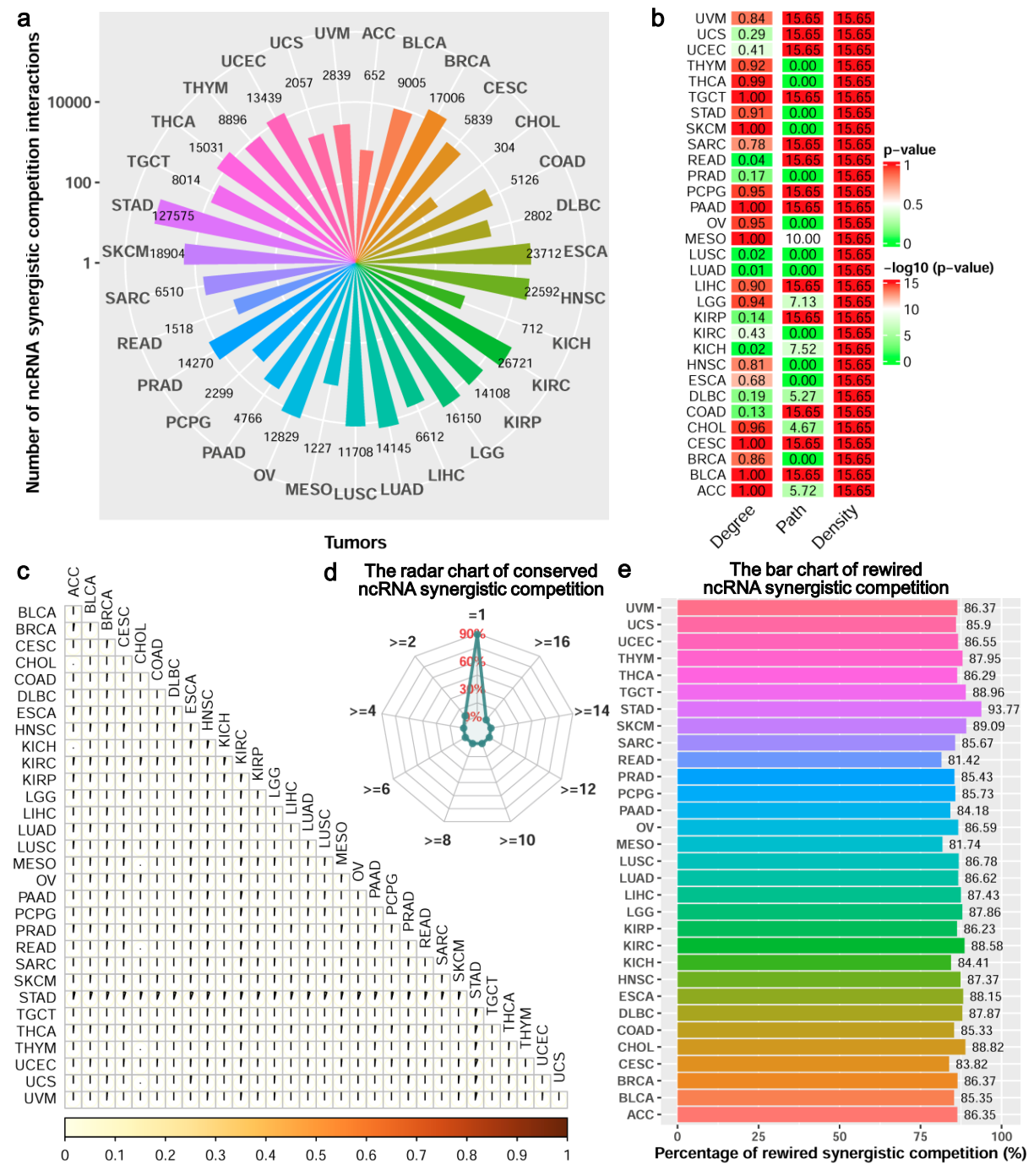
**S11 Data. Relationships between hub ncRNAs and immune cells across malignant tumors.**

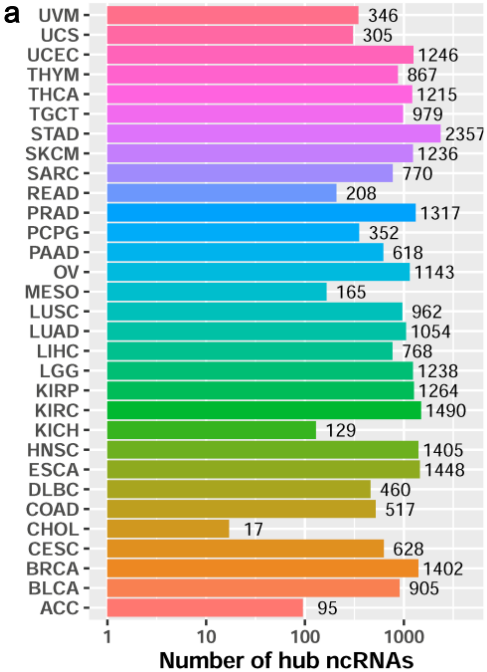
**S12 Data. Experimentally validated synergistic competition ncRNAs as biomarkers across malignant tumors.**

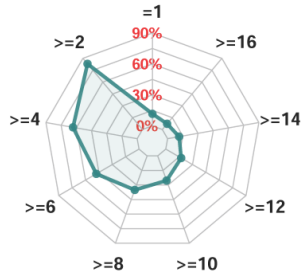
**S1 File. Supporting file.**


a
ceRNA network
mRNAs
ncRNA A
ncRNA B
Common mRNAs
Competition
Synergistic competition
b
Pan-cancer
LGG
HNSC
ESCA
THCA
PCPG
THYM
LUAD
BRCA
LUSC
LIHC
PAAD
STAD
CHOL
KICH
ACC
SARC
KIRC
KIRP
COAD
UCEC
READ
UVM
UCS
OV
BLCA
CESC
TGCT
PRAD
DLBC
MESO
SKCM
ncRNAs
Priori information
mRNAs
Pan-cancer tissues
miRNA-target interactions
Predicting ceRNA network
Predicted ncRNA-related ceRNA network
Tumor type 1
Tumor type 2
Tumor type C
ceRNA interactions
Inferring ncRNA synergistic competition network
Criterion I: Synergistic competition
Criterion II: Positive correlation
Criterion III: Sensitive correlation
Predicted ncRNA synergistic competition network
ncRNAs acting as ceRNAs
Synergistic competition


a
UVM
UCS
UCEC
THYM
THCA
TGCT
STAD
SKCM
SARC
READ
PRAD
PCPG
PAAD
OV
MESO
LUSC
LUAD
LIHC
LGG
KIRP
KIRC
KICH
HNSC
ESCA
DLBC
COAD
CHOL
CESC
BRCA
BLCA
ACC
0e+00
5e+05
1e+06
Number of ceRNA interactions
LncRNA
Pseudogene
b
LncRNA
Pseudogene
Degree
Path
Density
p−value
1
0.5
0
−log10 (p−value)
15
10
5
0
c
LncRNA
Pseudogene
ACC BLCA BRCA CESC CHOL COAD DLBC ESCA HNSC KICH KIRC KIRP LGG LIHC LUAD LUSC MESO OV PAAD PCPG PRAD READ SARC SKCM STAD TGCT THCA THYM UCEC UCS UVM
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
d
The radar chart of conserved ceRNA regulation
=1
>=2
>=4
>=6
>=8
>=10
>=12
>=14
>=16
90%
60%
30%
0%
LncRNA
Pseudogene
e
The ceRNA regulation in 31 malignant tumors
FGD5-AS1
HCG17
LY86-AS1
PDE4DIPP1
VPS35P1
NPIPB14P
ICE2P2
IKZF2
RPS27AP1
FPGS
PTEN
BCL2L12
HYAL4
KIR2DL1
CCNB2
LCE3D
ZDHHC19
NXPE4
CORIN
FZD3
SLC17A1
SH2D3C
CHD5
CDH26
FMO4
CD80
SERPINH1
SPHK1
VPS54
TGFBR2
E2F4
CHCHD2P9
PAICSP4
RPS4XP22
BTBD7
LINC00313
OGFR-AS1
ERI3-IT1
LINC00692
SEC24B-AS1
FAM215A
CIRBP-AS1
EIF4BP1
CNOT6LP1
SMG1P3
MYL6P3
NOC2LP2
VAV3
EDDM3A
CLN8
ENO1P1
CCDC191
USP37
EXOC7
FRS3
CD163
RHOU
PABPC4
SLC6A3
GOLT1B
BCL2L13
ELK3
CCNA1
TBC1D25
MED17
SESN2
CDK2AP1
SPANXA1
TTF2
CKAP2L
MBD4
TNRC18
UHRF2P1
ZNF292
LINC00862
TP53TG1
ACVR2B-AS1
CBR3-AS1
LINC00304
ACTG1P19
PMPCA
RPSAP16
RPS27AP9
IARS2P1
HNRNPCP8
TMSB4XP6
TRIM49C
RPL5P28
OPN1SW
TBX5
ORMDL3
TOMM20
ATXN2
NME3
CCNB1
POLR2D
SSNA1
RAB11FIP4
PRDM4
PHACTR4
TMPRSS11BNL
RPSAP42
PPT1
SGK1
SCAMP5
HDGFL3
ZNF669
DDX60
RPS3AP35
RPS3AP6
RBMS1P1
DNM2
ENTPD3-AS1
LINC01270
LINC01521
HCP5
HOXC-AS1
INHBA-AS1
ITPK1-AS1
PSMC1P1
NPM1P26
KDELC1P1
ZDHHC20P4
KRT18P14
PTP4A2P1
MRPS6P2
PTTG3P
MTND1P23
RPSAP43
MTND4P12
RPL17P49
HAUS1P2
RPL27AP5
SLC9B1P2
ASXL1
TRPC7
OR2H1
NAPA
PTP4A2
C1orf159
DCP1A
IFI27L1
ALPL
ZBTB10
LGR4
CPSF3
FNBP4
PSMA4
CSNK1G3
FBXW11
ALG1L
CEMIP
TENM3
TTC14
EXOC5
HNRNPM
ERMP1


a
Number of ncRNA synergistic competition interactions
10000
100
1
UVM 17006
ACC 5839
BLCA 304
BRCA 5126
CESC 2802
CHOL 23712
COAD 22592
DLBC 712
ESCA 26721
HNSC 14108
KICH 16150
KIRC 6612
KIRP 14145
LGG 11708
LIHC 1227
LUAD 12829
LUSC 4766
MESO 14270
OV 2299
PAAD 1518
PCPG 6510
PRAD 18904
READ 127575
SARC 8014
SKCM 15031
STAD 8896
TGCT 13439
THCA 2057
THYM 2839
UCEC 652
UCS 9005
Tumors
b
UVM 0.84 15.65 15.65
UCS 0.29 15.65 15.65
UCEC 0.41 15.65 15.65
THYM 0.92 0.00 15.65
THCA 0.99 0.00 15.65
TGCT 1.00 15.65 15.65
STAD 0.91 0.00 15.65
SKCM 1.00 0.00 15.65
SARC 0.78 15.65 15.65
READ 0.04 15.65 15.65
PRAD 0.17 0.00 15.65
PCPG 0.95 15.65 15.65
PAAD 1.00 15.65 15.65
OV 0.95 0.00 15.65
MESO 1.00 10.00 15.65
LUSC 0.02 0.00 15.65
LUAD 0.01 0.00 15.65
LIHC 0.90 15.65 15.65
LGG 0.94 7.13 15.65
KIRP 0.14 15.65 15.65
KIRC 0.43 0.00 15.65
KICH 0.02 7.52 15.65
HNSC 0.81 0.00 15.65
ESCA 0.68 0.00 15.65
DLBC 0.19 5.27 15.65
COAD 0.13 15.65 15.65
CHOL 0.96 4.67 15.65
CESC 1.00 15.65 15.65
BRCA 0.86 0.00 15.65
BLCA 1.00 15.65 15.65
ACC 1.00 5.72 15.65
Degree
Path
Density
p−value
1
0.5
0
−log10 (p−value)
15
10
5
0
c
ACC BLCA BRCA CESC CHOL COAD DLBC ESCA HNSC KICH KIRC KIRP LGG LIHC LUAD LUSC MESO OV PAAD PCPG PRAD READ SARC SKCM STAD TGCT THCA THYM UCEC UCS UVM
0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1
d
The radar chart of conserved ncRNA synergistic competition
=1
>=2
>=4
>=6
>=8
>=10
>=12
>=14
>=16
90%
60%
30%
0%
e
The bar chart of rewired ncRNA synergistic competition
UVM 86.37
UCS 85.9
UCEC 86.55
THYM 87.95
THCA 86.29
TGCT 88.96
STAD 93.77
SKCM 89.09
SARC 85.67
READ 81.42
PRAD 85.43
PCPG 85.73
PAAD 84.18
OV 86.59
MESO 81.74
LUSC 86.78
LUAD 86.62
LIHC 87.43
LGG 87.86
KIRP 86.23
KIRC 88.58
KICH 84.41
HNSC 87.37
ESCA 88.15
DLBC 87.87
COAD 85.33
CHOL 88.82
CESC 83.82
BRCA 86.37
BLCA 85.35
ACC 86.35
0 25 50 75 100
Percentage of rewired synergistic competition (%)



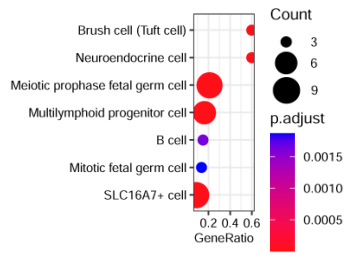



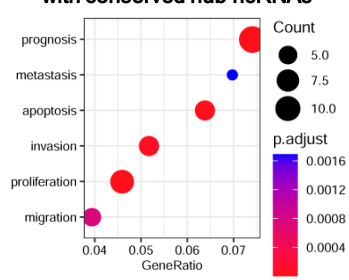

c Classification analysis of conserved hub ncRNAs

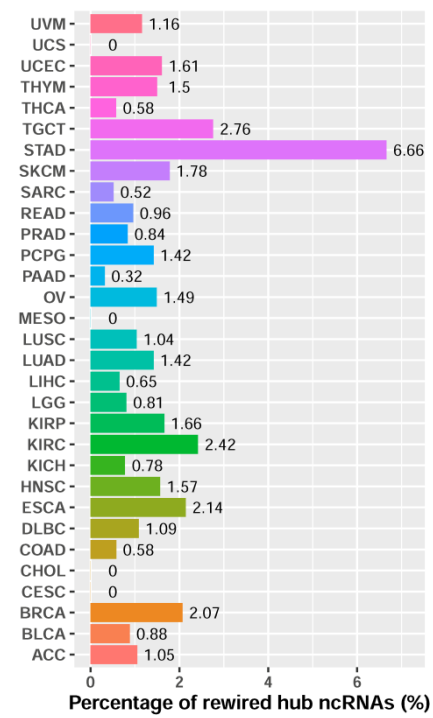
| Metrics | Conserved | Baseline |
|---|---|---|
| accuracy | **0.91** | 0.09 |
| average-precision | **0.95** | 0.17 |
| clp | **0.00** | 0.03 |
| coverage | **0.21** | 14.42 |
| F1 | **0.91** | 0.09 |
| hamming-loss | **0.01** | 0.06 |
| macro-AUC | **0.99** | 0.50 |
| macro-F1 | **0.89** | 0.01 |
| macro-precision | **0.89** | 0.00 |
| macro-recall | **0.90** | 0.03 |
| margin-loss | **0.21** | 14.42 |
| micro-AUC | **1.00** | 0.53 |
| micro-F1 | **0.91** | 0.09 |
| micro-precision | **0.90** | 0.09 |
| micro-recall | **0.92** | 0.09 |
| mlp | **0.00** | 0.97 |
| one-error | **0.09** | 0.91 |
| precision | **0.91** | 0.09 |
| ranking-loss | **0.01** | 0.48 |
| recall | **0.92** | 0.09 |
| subset-accuracy | **0.89** | 0.09 |
| wlp | **0.00** | 0.97 |

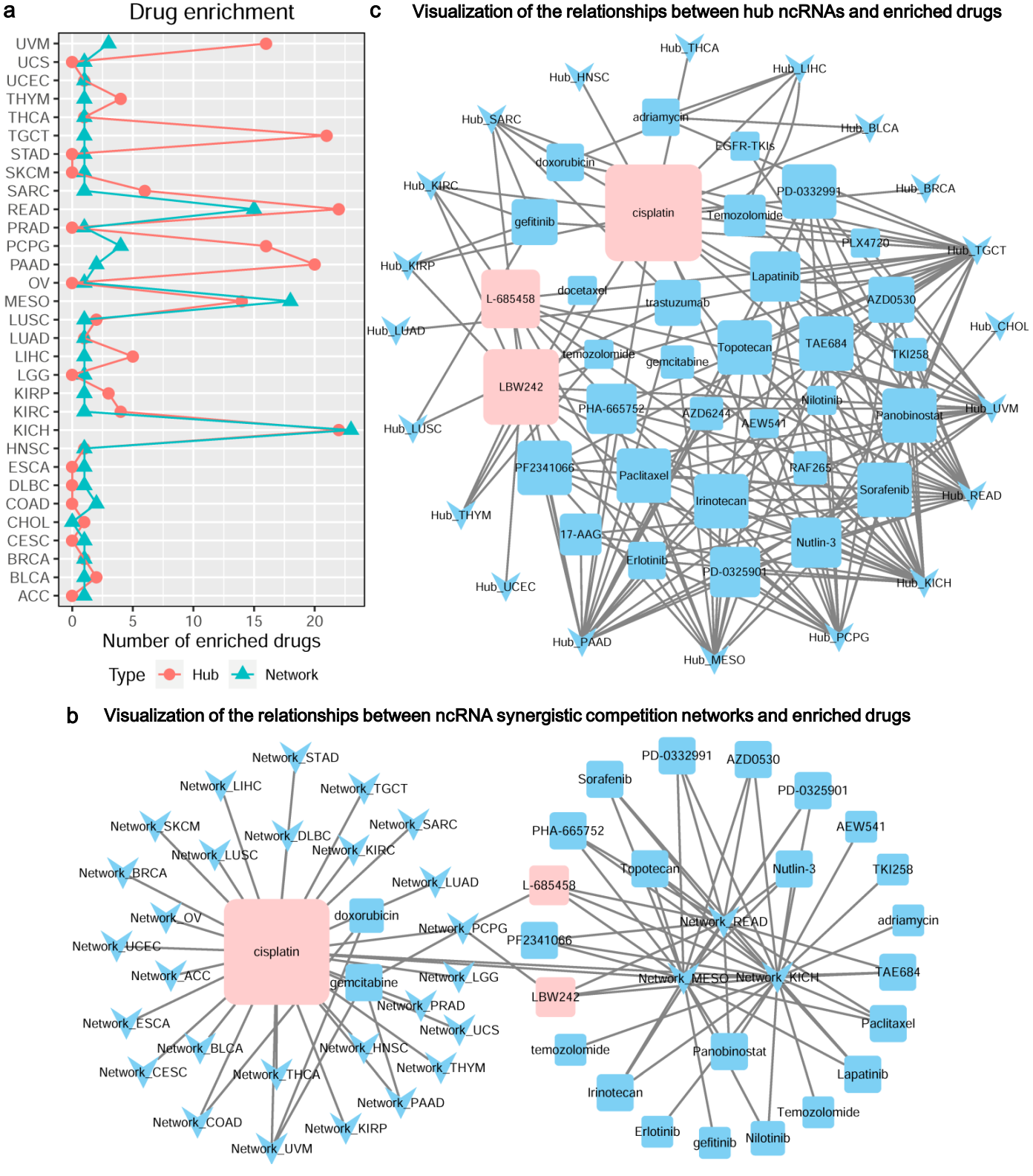
d Survival analysis of conserved hub ncRNAs

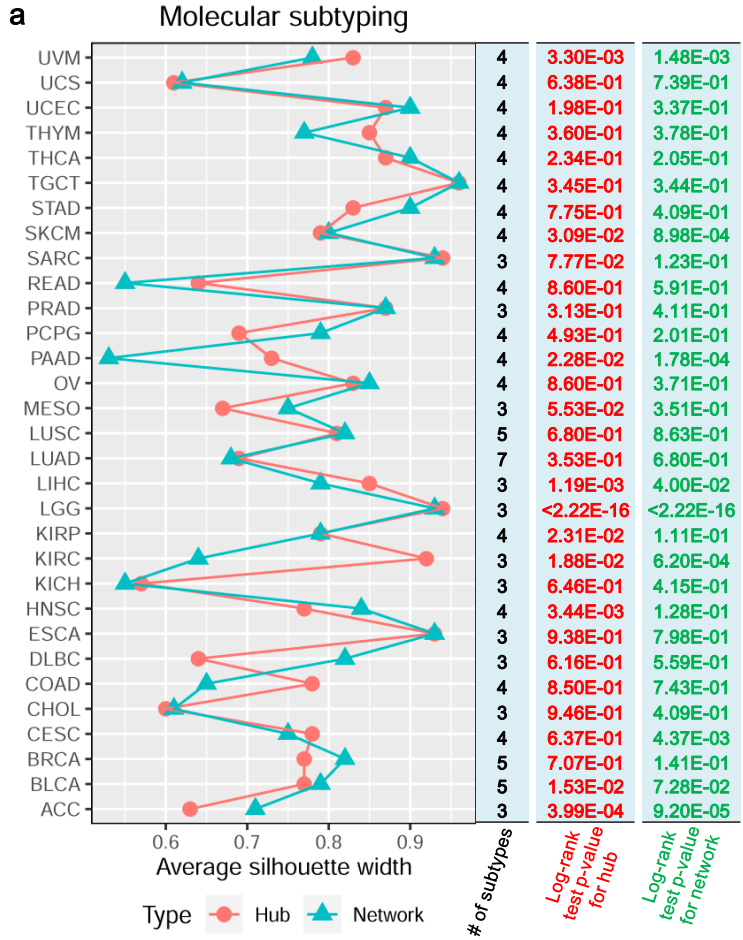
| Tumors | p-value | HR | HRlow95 | HRup95 |
|---|---|---|---|---|
| ACC | 5.33E-15 | 9.79 | 3.33 | 28.81 |
| BLCA | <2.22E-16 | 10.15 | 6.53 | 15.77 |
| BRCA | <2.22E-16 | 14.78 | 8.13 | 26.88 |
| CESC | <2.22E-16 | 8.80 | 3.99 | 19.42 |
| CHOL | 4.63E-09 | 9.31 | 2.08 | 41.74 |
| COAD | <2.22E-16 | 8.44 | 3.66 | 19.45 |
| DLBC | 6.51E-08 | 18.77 | 3.12 | 113.01 |
| ESCA | <2.22E-16 | 7.50 | 3.18 | 17.72 |
| HNSC | <2.22E-16 | 12.06 | 7.80 | 18.66 |
| KICH | 6.19E-07 | / | / | / |
| KIRC | <2.22E-16 | 15.26 | 7.05 | 33.00 |
| KIRP | <2.22E-16 | 11.18 | 4.04 | 30.94 |
| LGG | <2.22E-16 | 10.38 | 6.16 | 17.48 |
| LIHC | <2.22E-16 | 14.06 | 8.49 | 23.30 |
| LUAD | <2.22E-16 | 10.85 | 6.92 | 17.03 |
| LUSC | <2.22E-16 | 12.98 | 7.80 | 21.62 |
| MESO | <2.22E-16 | 7.31 | 2.99 | 17.90 |
| OV | <2.22E-16 | 7.49 | 4.75 | 11.81 |
| PAAD | <2.22E-16 | 7.69 | 3.83 | 15.44 |
| PCPG | 1.64E-05 | / | / | / |
| PRAD | 3.74E-10 | 23.38 | 5.30 | 103.27 |
| READ | 3.05E-10 | 10.00 | 2.38 | 42.11 |
| SARC | <2.22E-16 | 8.68 | 4.58 | 16.43 |
| SKCM | <2.22E-16 | 8.50 | 5.44 | 13.28 |
| STAD | <2.22E-16 | 10.95 | 6.16 | 19.47 |
| TGCT | 2.36E-03 | / | / | / |
| THCA | 7.27E-14 | / | / | / |
| THYM | 6.66E-09 | 20.61 | 3.23 | 131.69 |
| UCEC | <2.22E-16 | 64.84 | 35.96 | 116.91 |
| UCS | 1.11E-16 | 9.22 | 3.04 | 27.95 |
| UVM | 2.16E-14 | 11.05 | 3.19 | 38.26 |

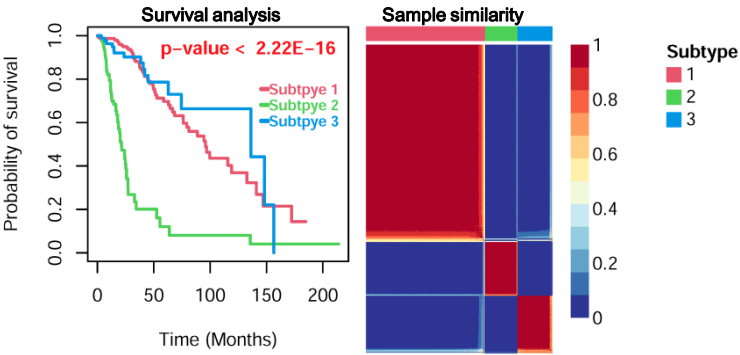


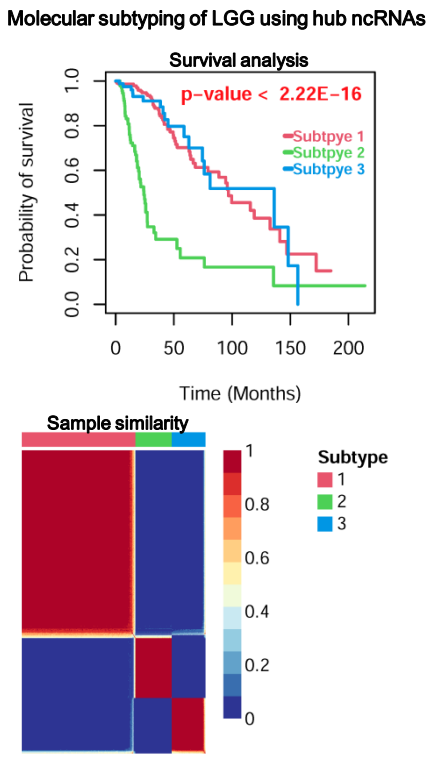




a
Drug enrichment
UVM
UCS
UCEC
THYM
THCA
TGCT
STAD
SKCM
SARC
READ
PRAD
PCPG
PAAD
OV
MESO
LUSC
LUAD
LIHC
LGG
KIRP
KIRC
KICH
HNSC
ESCA
DLBC
COAD
CHOL
CESC
BRCA
BLCA
ACC
0
5
10
15
20
Number of enriched drugs
Type
Hub
Network
c
Visualization of the relationships between hub ncRNAs and enriched drugs
Hub_THCA
Hub_HNSC
Hub_LIHC
Hub_SARC
Hub_BLCA
Hub_KIRC
Hub_BRCA
Hub_TGCT
Hub_KIRP
Hub_LUAD
Hub_CHOL
Hub_UVM
Hub_LUSC
Hub_READ
Hub_THYM
Hub_KICH
Hub_UCEC
Hub_PAAD
Hub_PCPG
Hub_MESO
cisplatin
L-685458
LBW242
adriamycin
doxorubicin
EGFR-TKIs
gefitinib
PD-0332991
Temozolomide
PLX4720
Lapatinib
docetaxel
trastuzumab
AZD0530
temozolomide
gemcitabine
Topotecan
TAE684
TKI258
PHA-665752
AZD6244
AEW541
Nilotinib
Panobinostat
PF2341066
Paclitaxel
RAF265
Sorafenib
Irinotecan
17-AAG
Nutlin-3
Erlotinib
PD-0325901
b
Visualization of the relationships between ncRNA synergistic competition networks and enriched drugs
Network_STAD
Network_LIHC
Network_TGCT
Network_SKCM
Network_DLBC
Network_SARC
Network_LUSC
Network_KIRC
Network_BRCA
Network_LUAD
Network_OV
Network_UCEC
Network_ACC
Network_ESCA
Network_BLCA
Network_CESC
Network_HNSC
Network_THCA
Network_THYM
Network_PAAD
Network_COAD
Network_KIRP
Network_UVM
Network_PCPG
Network_LGG
Network_PRAD
Network_UCS
cisplatin
doxorubicin
gemcitabine
PD-0332991
AZD0530
Sorafenib
PD-0325901
PHA-665752
AEW541
Topotecan
Nutlin-3
TKI258
L-685458
Network_READ
adriamycin
PF2341066
Network_MESO
Network_KICH
TAE684
LBW242
Paclitaxel
temozolomide
Panobinostat
Lapatinib
Irinotecan
Temozolomide
Erlotinib
gefitinib
Nilotinib



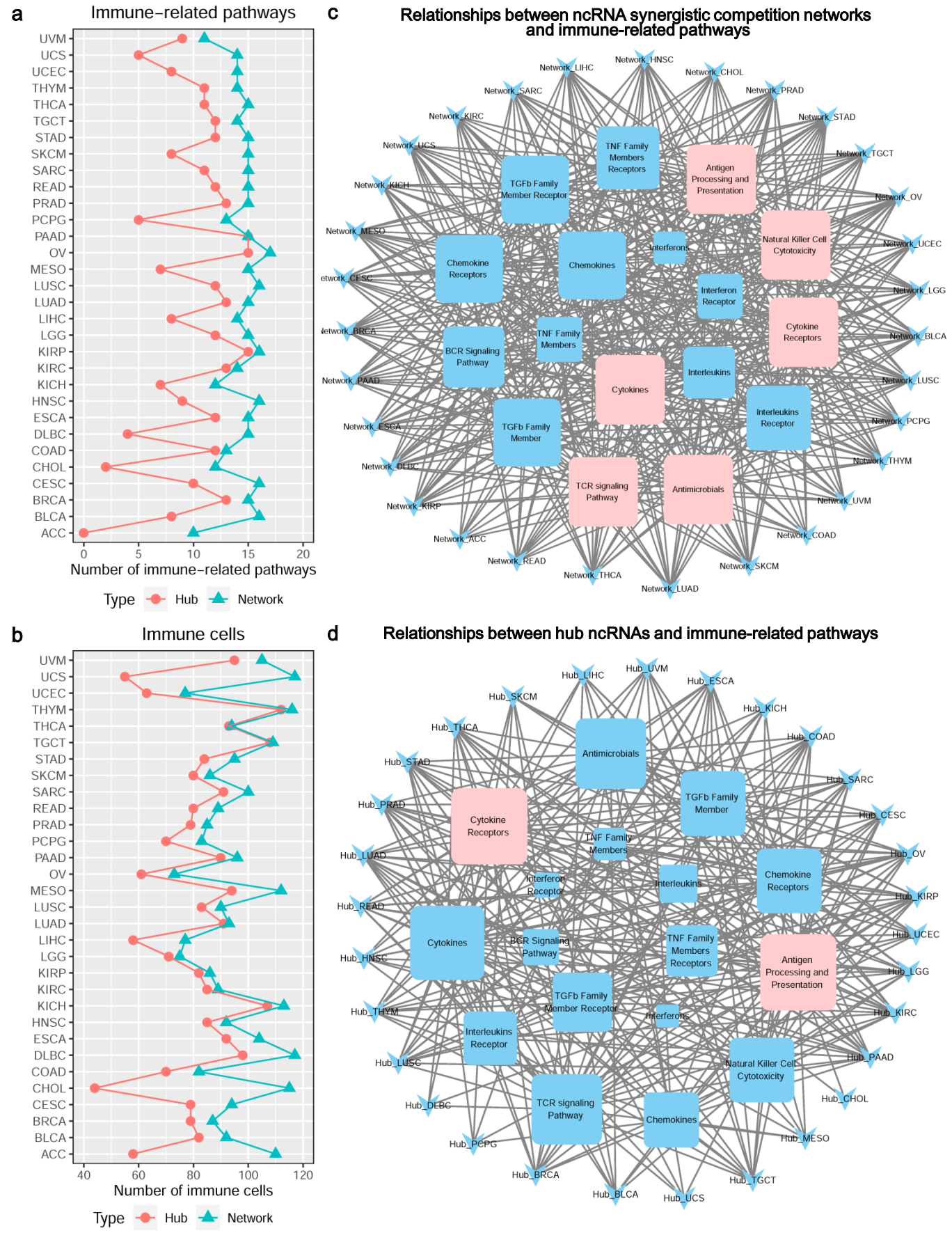



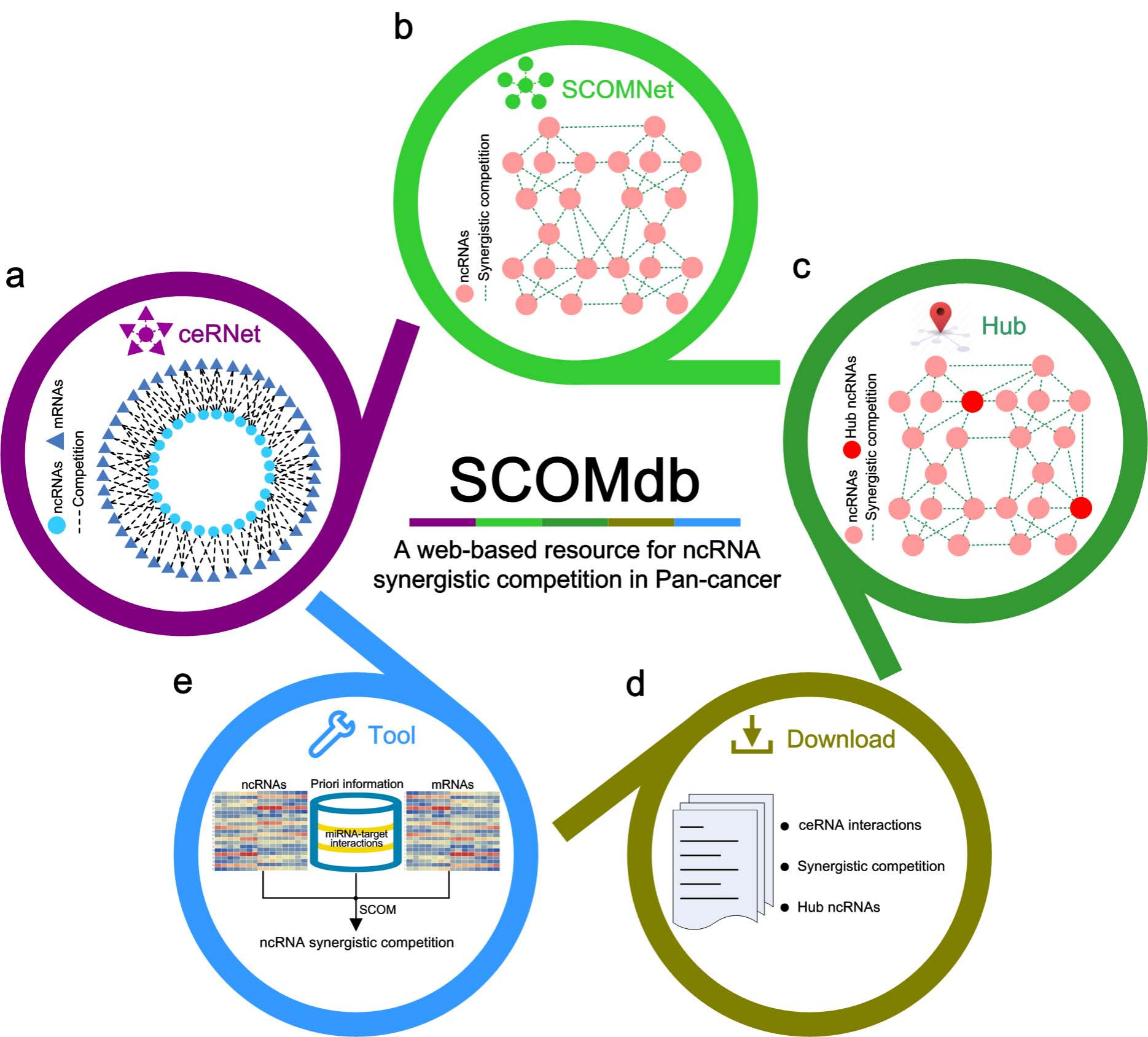




d

BRCA subtype classification

| Metrics | Network | Hub | Baseline |
|---|---|---|---|
| accuracy | **0.70** | **0.79** | 0.40 |
| average-precision | **0.83** | **0.89** | 0.60 |
| clp | **0.06** | **0.00** | 0.20 |
| coverage | **0.48** | **0.26** | 1.38 |
| F1 | **0.71** | **0.81** | 0.40 |
| hamming-loss | **0.12** | **0.08** | 0.24 |
| macro-AUC | **0.72** | **0.94** | 0.50 |
| macro-F1 | **0.59** | **0.74** | 0.11 |
| macro-precision | **0.61** | **0.75** | 0.08 |
| macro-recall | **0.62** | **0.75** | 0.20 |
| margin-loss | **0.48** | **0.26** | 1.38 |
| micro-AUC | **0.88** | **0.96** | 0.63 |
| micro-F1 | **0.70** | **0.80** | 0.40 |
| micro-precision | **0.68** | **0.77** | 0.40 |
| micro-recall | **0.72** | **0.83** | 0.40 |
| mlp | **0.24** | **0.02** | 0.80 |
| one-error | **0.30** | **0.20** | 0.60 |
| precision | **0.70** | **0.79** | 0.40 |
| ranking-loss | **0.12** | **0.06** | 0.34 |
| recall | **0.72** | **0.83** | 0.40 |
| subset-accuracy | **0.67** | **0.75** | 0.40 |
| wlp | **0.24** | **0.04** | 0.80 |


a
Immune−related pathways
UVM
UCS
UCEC
THYM
THCA
TGCT
STAD
SKCM
SARC
READ
PRAD
PCPG
PAAD
OV
MESO
LUSC
LUAD
LIHC
LGG
KIRP
KIRC
KICH
HNSC
ESCA
DLBC
COAD
CHOL
CESC
BRCA
BLCA
ACC
0
5
10
15
20
Number of immune−related pathways
Type
Hub
Network
b
Immune cells
40
60
80
100
120
Number of immune cells
Type
Hub
Network
c
Relationships between ncRNA synergistic competition networks and immune-related pathways
Network_HNSC
Network_LIHC
Network_CHOL
Network_SARC
Network_PRAD
Network_KIRC
Network_STAD
Network_UCS
Network_TGCT
Network_KICH
Network_OV
Network_MESO
Network_UCEC
Network_CESC
Network_LGG
Network_BRCA
Network_BLCA
Network_PAAD
Network_LUSC
Network_ESCA
Network_PCPG
Network_DLBC
Network_THYM
Network_KIRP
Network_UVM
Network_ACC
Network_COAD
Network_READ
Network_SKCM
Network_THCA
Network_LUAD
TNF Family Members Receptors
Antigen Processing and Presentation
TGFb Family Member Receptor
Natural Killer Cell Cytotoxicity
Interferons
Chemokine Receptors
Chemokines
Interferon Receptor
Cytokine Receptors
TNF Family Members
BCR Signaling Pathway
Interleukins
Cytokines
Interleukins Receptor
TGFb Family Member
TCR signaling Pathway
Antimicrobials
d
Relationships between hub ncRNAs and immune-related pathways
Hub_LIHC
Hub_UVM
Hub_ESCA
Hub_SKCM
Hub_KICH
Hub_THCA
Hub_COAD
Hub_STAD
Hub_SARC
Hub_PRAD
Hub_CESC
Hub_LUAD
Hub_OV
Hub_READ
Hub_KIRP
Hub_UCEC
Hub_HNSC
Hub_LGG
Hub_THYM
Hub_KIRC
Hub_LUSC
Hub_PAAD
Hub_DLBC
Hub_CHOL
Hub_PCPG
Hub_MESO
Hub_BRCA
Hub_TGCT
Hub_BLCA
Hub_UCS
Antimicrobials
TGFb Family Member
Cytokine Receptors
TNF Family Members
Chemokine Receptors
Interferon Receptor
Interleukins
Cytokines
BCR Signaling Pathway
TNF Family Members Receptors
Antigen Processing and Presentation
TGFb Family Member Receptor
Interferons
Interleukins Receptor
Natural Killer Cell Cytotoxicity
TCR signaling Pathway
Chemokines


a
ceRNet
mRNAs
ncRNAs
Competition
b
SCOMNet
ncRNAs
Synergistic competition
c
Hub
Hub ncRNAs
ncRNAs
Synergistic competition
SCOMdb
A web-based resource for ncRNA synergistic competition in Pan-cancer
d
Download
ceRNA interactions
Synergistic competition
Hub ncRNAs
e
Tool
ncRNAs
Priori information
mRNAs
miRNA-target interactions
SCOM
ncRNA synergistic competition