Integrated bioinformatic analysis of miR-654-5p target genes and experimental validation of functional analyses

Chuanyang Liu [1,†], Lu Min [1,†], Jingyu Kuang [1,†], Chu-shu Zhu [1], Jia-xin Ma [1], Xin-yuan Qiu [1], Xiaomin Wu [1], Lingyun Zhu [1,*]

[1] Department of Biology and Chemistry, College of Liberal Arts and Sciences, National University of Defense Technology, Changsha, Hunan, 410073, China
Lingyun Zhu
lingyunzhu@nudt.edu.cn

† These authors contributed equally to this paper

Abstract

Background: Lung cancer is one of the most malignant types of cancer worldwide. Recently, the pivotal role of miRNAs in carcinogenesis and tumor metastasis have been reported for their direct regulation of specific target genes. However, the precise roles of miR-654-5p in lung adenocarcinoma have been poorly investigated. In this study, we designed a series of closed-loop integrated bioinformatic analyses and *in vitro* experimental validation to explore the main function and regulation pattern of miR-654-5p, and elucidate its roles in lung adenocarcinoma metastasis and prognosis both *in silico* and *in vitro*.

Method: Data from The Cancer Genome Atlas (TCGA) were used to analyze the pan-cancer prognostic value of miR-654-5p. miRanda, TargetScan, RNA22 and miRWalks were utilized to screen the targets of miR-654-5p. Metascape and miRWalks were used to perform gene ontology analyses and multiple enrichment analyses. STRING database and Cytoscape were used to construct protein-protein network and identify hub genes. GEPIA 2.0 were used to perform pan-cancer expression and survival analyses of hub genes. The regulation between miRNAs and predicted genes were verified by qPCR, western blotting and dual-luciferase system. In addition, EMT hallmarks detection, cell proliferation assay and wound healing assay were used to demonstrate the predicted functions of miR-654-5p experimentally.

Results: We identified 275 potential targets of miR-654-5p and validated the direct regulatory effects of miR-654-5p on RNF8 *in vitro* as a proof of concept. Furthermore, we revealed the pivotal functions and disease association of miR-654-5p and experimentally validated the tumor suppressor roles of miR-645-5p that inhibited lung cancer cell epithelial-mesenchymal transition process, cell proliferation, and migration capacity. Among these 275 targets, 11 highly interconnected hub genes in lung adenocarcinoma were selected and studied through pan-cancer expression and pan-cancer survival analysis, in order to assess these targets' clinical value.

Conclusions: our research not only identified 275 hub target genes of miR-654-5p *in silico* and performed experimental validation but also revealed the tumor-suppressive roles of miR-654-5p in lung adenocarcinoma metastasis *in vitro* for the first time.

Keywords: lung cancer, bioinformatics, hub target genes, miR-654-5p, EMT

Background

Lung cancer is one of the most malignant types of cancer worldwide and is the primary cause of cancer-related deaths in most industrialized countries[1-3]. Non-small cell lung cancer (NSCLC) is the dominant type of lung cancer and accounts for approximately 80% of all lung cancer cases[4]. The principal cause of the high mortality among NSCLC patients is that NSCLC cells can easily metastasize to vital organs and finally lead to multiple organ failure[5,6]. As a consequence, the 5-year survival rate of NSCLC patients is still very low worldwide[7], highlighting the importance of mechanistic research about NSCLC and metastasis. As a complicated pathological phenomenon, the metastasis of cancer has been widely studied in recent years. However, the detailed molecular mechanism of lung adenocarcinoma (LUAD) metastasis remains unclear, which largely limits the development of target-driven drugs and therapeutics.

Recently, a class of short noncoding RNAs, microRNAs (miRNAs), have been proven to play important roles in cancer metastasis via post-transcriptional regulation of their target genes' expression[8-11]. For example, miR-654-5p has been reported to attenuate breast cancer progression by targeting EPSTI1[12]. In addition, miR-654-5p can target GRAP to promote proliferation, metastasis, and chemoresistance of oral squamous cell carcinoma through Ras/MAPK signaling[13]. These studies provide some specific explanation for how miR-654-5p regulates some specific biological processes in cancers. However, in a larger landscape, the common functions of miR-654-5p and how it regulates cellular pathway networks or biological processes have been poorly investigated and assessed. Thus, the identification and functional analysis of miR-654-5p-targeting hub genes would provide a more fundamental insight into the mechanism underlying miR-654-5p-induced downstream signaling transduction or phenotype alterations.

Here, we aimed to elucidate the main functions and regulation pattern of miR-654-5p from all angles, and thus we designed a series of experiments integrating bioinformatics analysis and *in vitro* validation experiments (Fig. 1). Concretely, we first identified 275 potential targets of miR-654-5p by four miRNA-targets prediction tools, and then experimentally proved the direct regulatory effect of miR-654-5p on RNF8, a target with high binding site accessibility. Then, GO annotation and multiple enrichment analyses were performed by Metascape, which revealed that miR-654-5p targets were significantly associated with cell growth, cell-cell adhesion, and epithelial-mesenchymal transition. Therefore, we further designed *in vitro* cell proliferation and migration experiments that validated that miR-645-5p could act as a tumor suppressor to inhibit the lung cancer cell EMT process, cell proliferation, and migration capacity. These *in vitro* results not only proved our bioinformatic functional enrichments but also exhibited novel biological functions of miR-654-5p.

To reveal more associations between miR-654-5p and different cancers, we screened hub target genes by integrating results from differential expression analysis and differential survival rate analysis, which identified many cancer types-specific miR-654-5p targets-regulation axes. By constructing a protein-protein network, we revealed target genes' potential interactions and identified five interconnected clusters. Finally, 11 highly interconnected hub genes in lung adenocarcinoma were selected to perform pan-cancer expression and survival analysis to assess their pan-cancer expression pattern and clinical value. In summary, our research not only identified the hub target genes of miR-654-5p *in silico* and performed corresponding experimental validation but also revealed novel anti-oncogenic roles of miR-654-5p in lung adenocarcinoma metastasis *in vitro* for the first time.

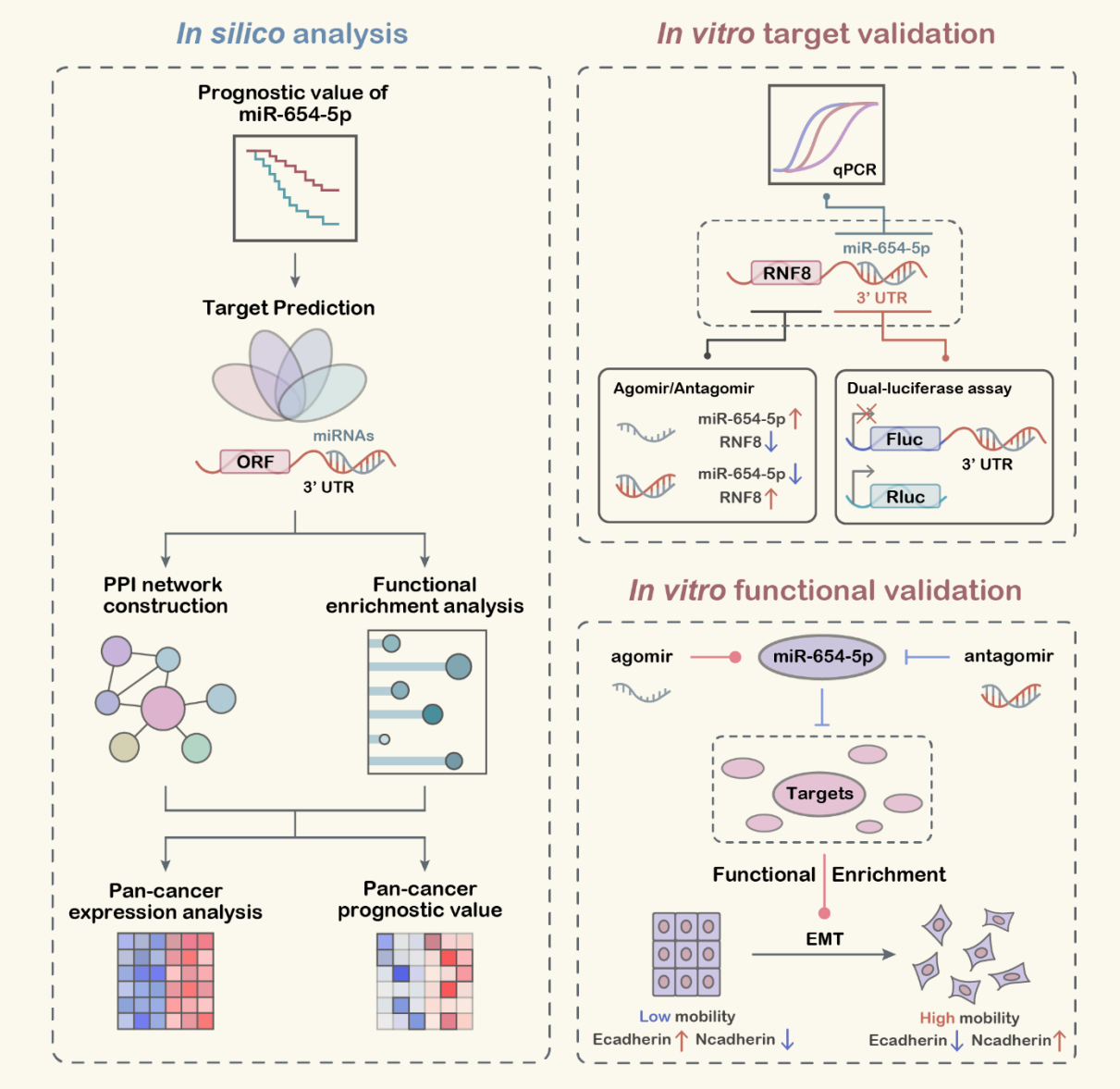

Figure 1. Graphical representation of the workflow in the present study.

1 Methods

1.1 Evaluation of the prognostic value of miR-654

Pan-cancer overall survival analysis were performed based on pan-cancer TCGA miRNA database downloaded from Kaplan-Meier Plotter (http://kmplot.com/analysis/)[14]. The R package "survival" and "survminer" was used for survival analysis and visualization with 50% cutoff of miR-654 expression. A logrank $p<0.05$ were considered as significant. The data of miR-654 and clinical pathological features were downloaded from TCGA database MEXPRESS[15].

1.2 Integrated target prediction of miR-654-5p target genes

MiRWalk 3.0 (http:// http://mirwalk.umm.uni-heidelberg.de/), an easy-to-access online tools including predicted miRNAs-mRNA data obtained with a machine learning algorithm, were designed to provide promising and mostly up-to-date information for miRNA-related researches[16]. We analyzed the potential miR-654-5p target genes with miRWalks and 3 other highly recognizable miRNA-target prediction tools (miRanda: http://www.microrna.org/microrna/getMirnaForm.do, miRDB: http://mirdb.org/ and Targetscan7.2: http://www.targetscan.org/vert_72/). The relationships between these targets' sets were analyzed and visualized by R package "VennDiagram".

1.3 Bioinformatic exploration of the functions of miR-654-5p target genes

1.3.1 Functional enrichment analysis

The 275 overlapping genes were then submitted to the Metascape (metascape.org/gp/index.html), a user-friendly tool for gene annotation and analysis resource that assists biologists to make sense of one or multiple gene lists[17], for gene ontology (GO) annotation and enrichment analyses including KEGG (Kyoto Encyclopedia of Genes and Genomes) pathway. $P<0.05$ was considered as significant. Furthermore, Hallmark ($p<0.05$), KEGG functional sets ($p<0.05$), Oncogenic signatures ($p<0.05$), Reactome ($p<0.001$), BioCarta ($p<0.05$) and Canonical pathway ($p<0.01$) enrichment analysis enrichment analysis were also performed by KOBAS database. In addition, the predicted targets of miR-654-5p by miRWalks 3.0 (with score>0.95) were also submit to its own GO annotation and KEGG pathway enrichment analysis to avoid the key information missing after only considering the intersection of predicted targets. R package "ggplot2" were used to visualized the results of GO annotation and KEGG pathway enrichment analysis.

1.3.2 Protein-protein interaction analysis

Protein-protein interaction (PPI) network of overlapping genes was construct by the Search Tool for the Retrieval of Interacting Genes (STRING, https://string-db.org/). In present study, high confidence (minimum required interaction score>0.700) was set as the selection criterion of constructing the network, all disconnected nodes were excluded from the network. The list of PPI pairs was downloaded for further analysis and visualized by Cytoscape software (version 3.7.1). Molecular Complex Detection (MCODE) plugin in Cytoscape was utilized to find potential cluster in the PPI network based on topology, which may help identify the most likely key target genes for miR-654-5p. The degree cut-off value to 2 and the node score cut-off to 0.2 were set in the MCODE process. Genes with those

differentially expressed in lung adenocarcinoma (LUAD) and highly connected (Degree >2) were selected as hub genes. Genes with differential expression (DEGs) and differential survival rate (DSGs) in various cancers were obtained using online tools Gene Expression Profiling Interactive Analysis (GEPIA) tool (http://gepia.cancer-pku.cn/). DEGs screening were based on R package limma with $|Log_2FC|>1$ and p-value<0.05. top 500 DSGs were selected for further analysis.

1.4 Expression pattern and prognostic value of hub target genes

1.4.1 Pan-cancer expression of hub genes

In order to explore the expression pattern of hub genes in various cancers, Gene Expression Profiling Interactive Analysis (GEPIA) tool (http://gepia.cancer-pku.cn/) were utilized the to compare the expression of overlapping genes in cancers and their corresponding normal tissue. GEPIA integrated mRNA sequencing data from TCGA and the Genotype-Tissue Expression (GTEx) project, providing customizable functionalities for differential expression analysis, profiling plotting, patient survival analysis, and so on[18,19]. The pan-cancer expression of hub genes was shown as heatmap using R package "pheatmap".

1.4.2 Pan-cancer prognostic value of hub genes

In order to assess the prognostic value of hub genes in lung cancer, we firstly analyzed the overall survival rate of these genes in lung cancer based on TCGA database (Tumor: 483 samples; Normal: 347 samples) downloaded from Kaplan-Meier Plotter (http://kmplot.com/analysis/) [14,20]. To further assess the prognostic value of these genes, Gene Expression Profiling Interactive Analysis (GEPIA) tool (http://gepia.cancer-pku.cn/), including integrated TCGA mRNA sequencing data and the GTEx, were also used (with FDR P-value adjustment, 0.05 significance level and Median group cut-off) to calculate patient overall survival rate (OS) and relapse-free survival rate (RFS) [18,19]. The results were shown in form of heatmap with colors of cells showing log_{10}(HR) and the frame meaning significance.

1.5 *In vitro* proof of concept for bioinformatic analyses

1.5.1 Cell culture

Lung cancer cell lines A549, H1299 and human embryonic kidney cell line 293T were purchased from American Type Culture Collection (ATCC, Manassas, VA), and cultured in RPMI 1640 Medium (Hyclone) supplemented with 10% fetal bovine serum (FBS; GIBCO, Gaithersburg, MD, USA) and 100 U/ml penicillin and streptomycin (P/S; Hyclone). Cells were contained in a 5% CO_2 incubator at 37°C.

1.5.2 Overexpression or knockdown of miRNA and RNF8

in vitro overexpression and knockdown of miR-654-5p were performed by liposome-mediated miRNA Agomir and Antagomir transfection according to the . Chemically modified hsa-miR-654-5p mimics and hsa-miR-654-5p inhibitor, with their negative control Stable N.C and inhibitor N.C respectively, were purchased from GenePharma. RNF8 siRNA (si-RNF8-1, 5′-GGACAAUUAUGGACAACAA-3′) was purchased from GenePharma. A549, H1299 and HEK293T cells were transfected using

Lipofectamine 2000 (Invitrogen) according to the manufacturer's directions with the working concentration of Antagomir/Agomir 20nM, siRNA 30nM.

1.5.3 RNA extraction and quantitative real-time PCR

Total RNAs were extracted using the TRIzol agent (Ambion) according to the instruction of the manufacturer. Reverse transcription of RNA and quantitative real-time PCR was performed using the Hairpin-it^{TM} miRNAs qPCR Quantitation Assay Kit (GenePharma) according to the manufacturer's instructions. Quantitative RT-PCR was performed in a Roche 480 real-time PCR system. The $2^{-\Delta\Delta Ct}$ method was used to evaluate the miR-654-5p gene expression after normalization for expression of the endogenous controls U6 (U6 non-coding small nuclear RNA). All primers for miR-654-5p and the U6 genes were synthesized and approved by GenePharma. Each experiment was repeated at least three times.

1.5.4 Western Blotting

Total proteins of cells were extracted using RIPA lysis buffer (Beyotime) with Protease Inhibitor (Roche) and Phosphatase Inhibitor (Roche). Protein samples were separated in sodium dodecyl sulfate (SDS)-PAGE and transferred to polyvinylidene fluoride (PVDF) filter membranes (Millipore, USA) for immune-hybridization. After 1 hour blocking in PBST (phosphate buffered saline containing 0.05% Tween-20 and 5% non-fat milk powder), the membranes were incubated with one of the following primary antibodies with corresponding concentration: RNF8 (Santa Cruz Biotech, 1:500), EMT kit (Cell Signaling Technology, 1:2000), beta-actin (Santa Cruz Biotech,1:4000), Secondary antibodies were Horseradish peroxidase (HRP)-conjugated anti-mouse IgG (ZB-2305,ZSGB-Bio,1:4000) or anti-Rabbit IgG(Fc) (ZB-2301, ZSGB-Bio, 1:4000). Subsequently, band visualization was performed by electro-chemiluminescence (ECL) and detected by Digit imaging system (Thermo, Japan), the gray level of the bands was quantitated by ImageJ software.

1.5.5 *In vitro* cell proliferation assays

To perform cell proliferation assays, cells were counted and plated in the well of 96-well plate (1500 cells per well) 24h after transfection of chemically modified oligonucleotides or siRNA. the cell proliferation ability was determined using the Cell Counting Kit-8(CCK8) Assay Kit (Dojindo Corp, Japan) according to the manufacturer's protocol: After the 0/24/48/72h proliferation of cells, the kit reagent dissolved with RPIM1640 Medium to prepare a 10% working reagent. The original medium was removed and 110ul working reagent was added to each well. After 2 hours incubation in the 37℃ incubator. The absorbance was measured at 450nm to calculate the number of cells, the proliferation assays were performed three independent times.

1.5.6 *In vitro* wound healing assays

Cell migration assays were performed using a 6-well plate. 24h after transfection of synthesized miRNA oligonucleotide or siRNA, cells should reach ~70-80% confluence. 20μL pipette tip were utilized to scratch the monolayer cells after washing by PBS. Gently wash the well twice with PBS again to remove the detached cells and replenish the well with fresh complete medium. After 24 hours

incubation. Cell migration was imaged by microscope and quantified by imageJ software. The assay was performed three independent times.

1.5.7 Statistical analysis

Statistical analysis was conducted using the GraphPad Prism (Version 8.0; SPSS Inc.). All results were presented as the mean ± standard error of the mean (SEM). Student *t*-test and One-way ANOVA was performed to compare the differences between treated groups relative to their paired controls. p values are indicated in the text and figures above the two groups compared and $p<0.05$ (denoted by asterisks) was considered as statistically significant.

2 Results

2.1 *In silico* prognostic analysis of miR-654-5p in cancers

To assess the prognostic value of miR-654-5p in various cancers, we performed pan-cancer survival analysis based on the pan-cancer TCGA miRNA database[21]. The results showed that high abundance of miR-654 was significantly associated with poor prognosis of bladder cancer (BCa, Fig. 2 A, log-rank $p=7.8\times10^{-4}$), head-neck squamous cell carcinoma (HNSC, Fig. 2 B, log rank $p=1.4\times10^{-4}$), kidney renal papillary cell carcinoma (KIRP, Fig. 2 C, log rank $p=3.6\times10^{-4}$), ovarian cancer (OV, Fig. 2 D, log rank p=0.018), stomach adenocarcinoma (STAD, Fig. 2 F, log rank p=0.027), and thyroid carcinoma (THCA, Fig. 2 G, log rank p=0.0029). In rectum adenocarcinoma (READ), miR-654 abundance was positively correlated with the poor prognosis (Fig. 2 E, log rank p=0.036). In other cancers, however, the relation of miR-654 and survival rate were not significant (Supplementary Fig. S1 B, log rank p>0.05). These pan-cancer survival results demonstrated that miR-654-5p abundance might be associated with the progression of these cancers. among clinical pathological features of LUAD, we found miR-654-5p were significantly associated with histological type (Fig. 2 I, log rank p=0.005). In summary, miR-654-5p could be a promising prognostic marker for patients' survival.

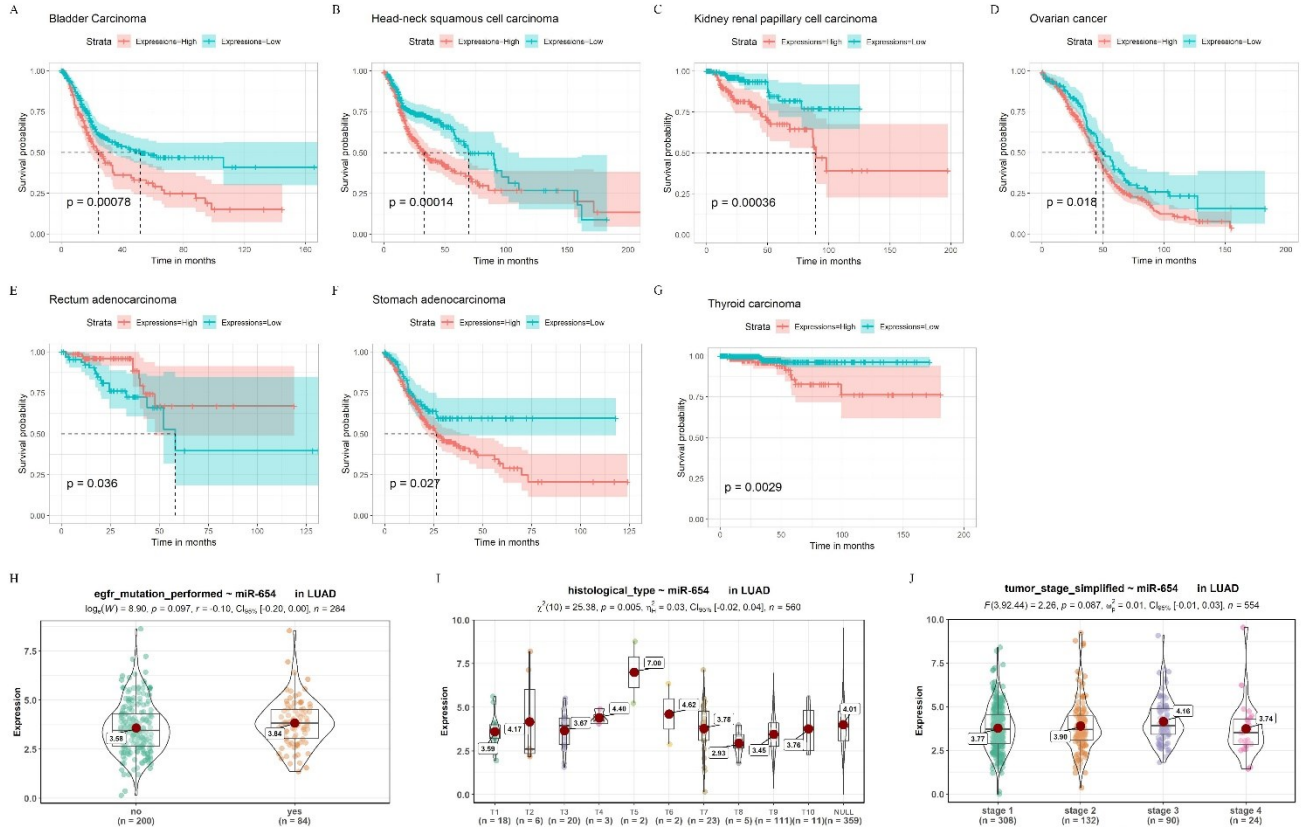

Figure 2. Survival analysis of miR-654. (A-G) Overall survival (OS) of hsa-miR-654 based on TCGA pan-cancer database[14]. (H-J) the correlation between miR-654 and clinical pathological features including EGFR mutation performed (H), histological type (I) and simplified tumor stage (J) based on 861 TCGA LUAD samples [15].

2.2 Targets prediction and *in vitro* validation of miR-654-5p

2.2.1 Integrated targets prediction of miR-654-5p

The common regulation pattern of miRNAs is to bind to the 3' untranslated region (3'UTR) and inhibit the translation of target genes[10,22,23]. Thus, the first step to understanding the core function of a miRNA is to identify its targets. miRWalks 3.0 and three other highly promising miRNA-target prediction tools were utilized to predict the targets of miR-654-5p. Subsequently, the intersection of these four predicted targets' sets were visualized by R (Fig. 3 A)[24,25]. We found 275 overlapping genes were predicted by all four tools (Fig. 3 A and B), indicating that these genes should be promising targets of miR-654-5p. Among these genes, those with high target sites accessibility, including RNF8, CYP4A11, and WASF2, are shown in Fig B, indicating that binding to these gene sites were common and accessible for miR-654-5p in cells, which could easily be verified *in vitro*.

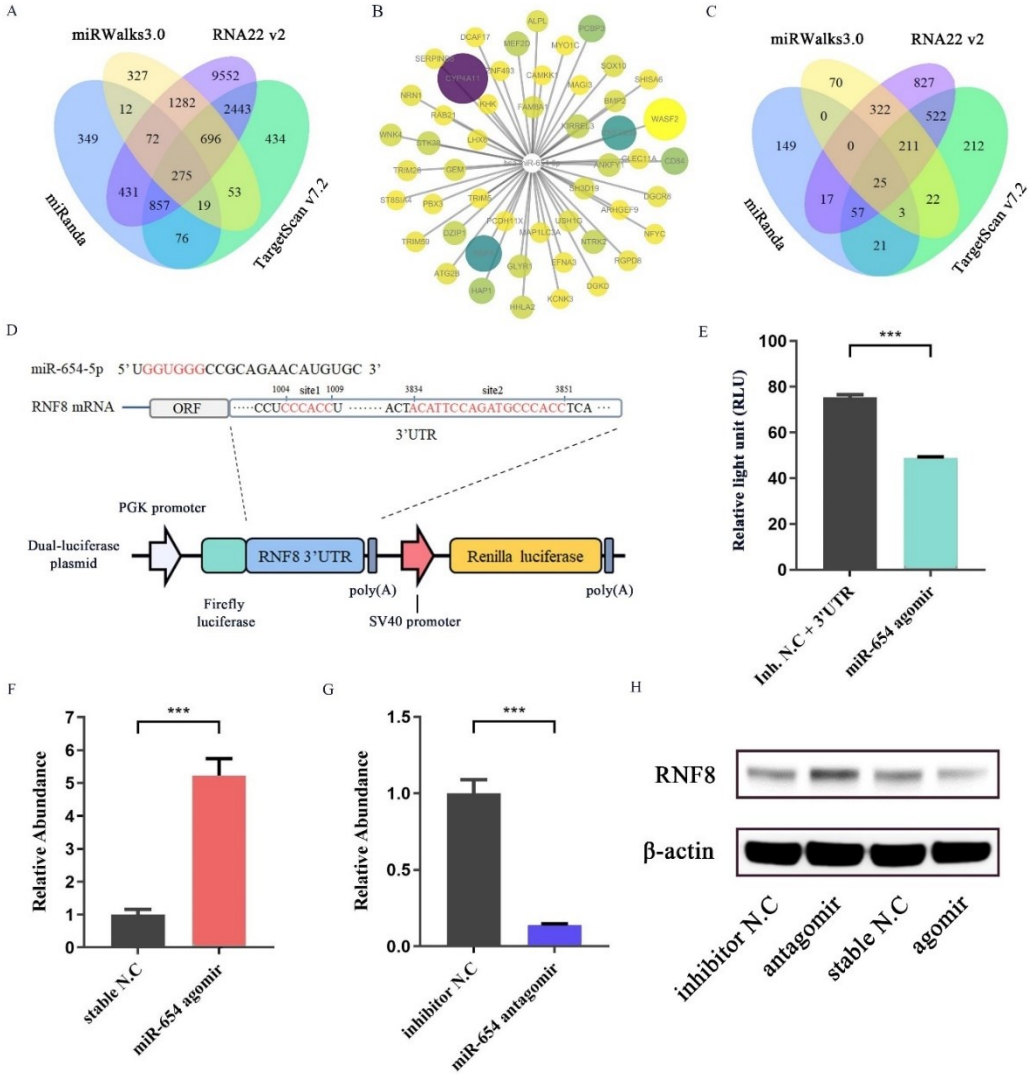

Figure 3. Potential targets predicted by multiple tools. (A) Four promising miRNA-targets prediction tools including miRanda, TargetScan7.2, RNA22 v2.0, and miRWalks v3.0 were used to find the targets of miR-654-5p. (B) miR-654-5p-targets interaction network was visualized using Cytoscape 3.7.1, the deeper color and larger sizes

of nodes indicate the high accessibility of each gene. (C) Four tools were utilized to reverse miRNAs prediction using RNF8 3'UTR sequencing. (D) the binding between miR-654-5p and RNF8 3'UTR and the design of dual-luciferase plasmid. (E) 48 hours after the liposome-mediated transfection of miR-654-5p agomir/stable NC and designed dual luciferase plasmid, dual-luciferase assay was performed. Relative light unit (RLU) was obtained by the ratio of Firefly fluorescence/Renilla fluorescence (F and G) Real-time PCR was performed to detect the abundance of miR-654-5p in A549 cells treated with agomir/antagomir. (H) Western blot assay was performed to measure the expression of RNF8 in treated cells. miR-654-5p RNA levels are expressed as the mean ± SEM of four different experiments normalized to U6 abundance. *p<0.05, **p<0.01, ***p<0.001 vs. control.

2.2.2 miR-654-5p/RNF8 axis: a case of *in vitro* validation of predicted targets

Among 275 targets predicted by the four tools, RNF8, a predicted target hub gene that has a high target sites accessibility, has recently been proven to have a strong relation to breast cancer[26-29]. We therefore chose RNF8 as an *in vitro* proof of concept for our bioinformatic targets analysis of miR-654-5p. In addition, along with forward prediction using the sequence of miR-654-5p, reverse miRNAs prediction using RNF8 3'UTR also indicated that miR-654-5p is one of the 25 miRNAs predicted by all four tools (Fig. 3 C), which further demonstrated that the axis of miR-654-5p-RNF8 might be pivotal to the roles of both miR-654-5p and RNF8 in the cell regulation network.

To validate the regulation of miR-654-5p on RNF8 *in vitro*, the potential binding sites of miR-654-5p in the RNF8 3'UTR were screened (Fig. 3 D). Then, we cloned the full-length 3'UTR sequence of RNF8 and linked it behind the 3' end of the coding sequence of luciferase. To simulate the natural transcriptional inhibition of miR-654-5p on RNF8, the frequently used 293T cell line was transfected with the RNF8 3'UTR-inserted dual-luciferase system and miR-654-5p mimics (agomir), which could reveal any difference of liposome-mediated transfection. The results showed that the expression of luciferase was decreased significantly compared to the control group to which miR-654-5p mimics were not introduced (Fig. 3 E). This result demonstrated that, based on the regulation pattern of the miRNAs, miR-654-5p could regulate upstream luciferase expression via directly targeting the inserted RNF8 3'UTR.

To further prove that miR-654-5p inhibits the expression of RNF8 *in vitro*, lung cancer cells A549 were transfected with miR-654-5p agomir or antagomir. The results showed that miR-654-5p's abundance in A549 was upregulated 5.23-fold (Fig. 3 F, $p<0.001$) by agomir-transfection and the RNF8 protein level was dramatically downregulated accordingly (Fig. 9 H, right) in the antagomir-group. In contrast, when endogenous miR-654-5p was downregulated by transient transfection of chemically modified miR-654-5p antagomir (downmodulated approximately 0.86-fold, $p<0.001$, compared with inhibitor NC, Fig. 3 G, left), the RNF8 protein level was increased (Fig. 9 H). Combined with the results of the dual-luciferase assay that miR-654-5p could bind RNF8 at its 3'UTR to induce post-transcriptional silencing, we concluded that miR-654-5p can directly regulate RNF8 in lung cancer cells, which verified the axis of miR-654-5p-RNF8.

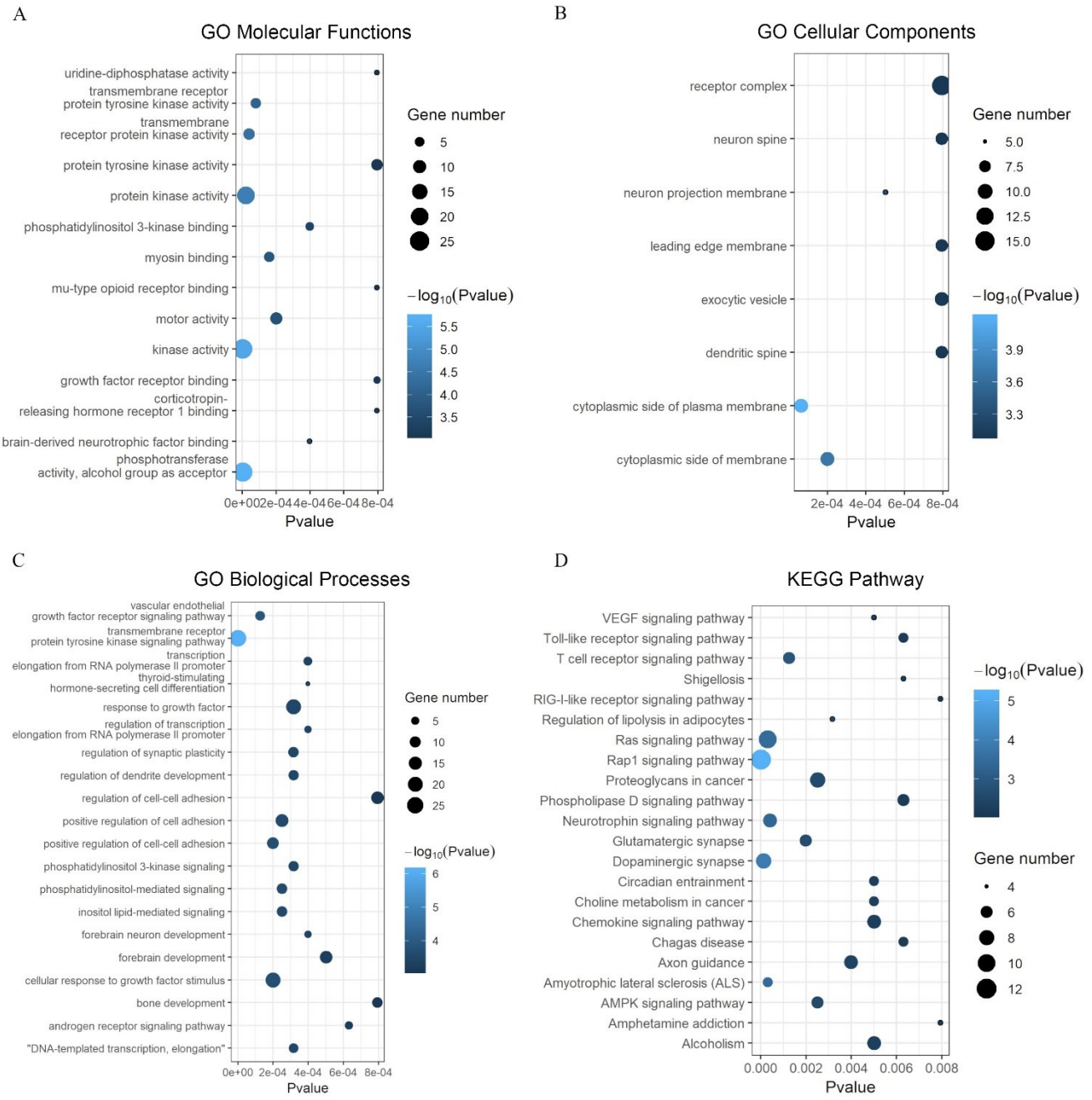

Figure 4. Gene ontology and KEGG pathway enrichment analysis of 275 overlapping predicted targets of miR-654-5p. Gene ontology (GO) annotation analysis and KEGG pathway enrichment analysis results were performed by Metascape and visualized by R package "ggplot2" based on 275 overlapping genes. Each bubble represents a term, and its size represents the counts of involved genes. Lighter colors of bubbles indicate smaller P values. (A) Enriched terms of GO molecular functions (MFs, $p<0.001$). (B) Enriched terms of GO biological processes (BPs, $p<0.001$). (C) Enriched terms of GO cellular compounds (CCs, $p<0.001$). (D) Enriched terms KEGG pathway ($p<0.01$).

2.3 Functional enrichment analysis of the predicted target genes and *in vitro* validation

2.3.1 GO annotation and KEGG enrichment analyses

To obtain a better understanding of regulation pattern and functions of miR-654-5p at a cellular level, for the predicted 275 target genes of miR-654-5p we performed GO annotation and KEGG pathway enrichment using the web-based functional enrichment tool Metascape[17]. Terms with P values of <0.01 were considered as significant and visualized by R. As shown in the results, molecular functions (MFs) terms such as kinase activity, protein kinase activity, and phosphotransferase activity were significantly enriched (Fig. 4 A). In GO biological processes (BPs) enrichment, these 275 targets were mainly enriched by the VEGFR signaling pathway, RTK signaling pathway and cellular response to growth factor stimulus, indicating miR-654-5p might be crucial in RTK-related biological processes such as the regulation of cell proliferation. Multiple cell adhesion-related items such as regulation of cell-cell adhesion were also enriched, demonstrating miR-654-5p is likely to regulate cell adhesion (Fig. 4 C). For cellular components (CCs), target genes were commonly enriched by cytoplasmic side of plasma membrane, cytoplasmic side of membrane, and receptor complex (Fig. 4 B). Furthermore, RTK pathway-related pathways such as VEGF and Ras signaling were enriched, thereby corroborating the results of the GO BPs (Fig. 4 D).

To further explore the correlation between miR-654-5p and diseases and pathways, Hallmark gene sets, oncogenic signatures, Reactome, and KEGG functional sets were also performed by Metascape based on the 275 genes, and the results showed that in the KEGG functional sets, PI3K-Akt signaling and MAPK (p38 and JNK) signaling were enriched (Supplementary Fig. S2). Similar results were also seen in the BioCarta gene set and Canonical pathways enrichment (Supplementary Fig. S3 E and F), and these results indicate miR-654-5p might be involved in the regulation of cell proliferation, migration, apoptosis etc. For Reactome, RTK-related items like signaling by the RTK, VEGF and VEGFA-VEGFR2 pathways were enriched as expected (Supplementary Fig. S3 D), further proving the regulatory role of miR-654-5p in cell proliferation. For Hallmarks enrichment, items about interferon response, inflammation and epithelial-mesenchymal transition were enriched (Supplementary Fig. S3 A).

In oncogenic signature, TGF-β up, STK33 down, ESC v6.5 up, CAMP up and AKT up/mTOR down were most significantly enriched (Supplementary Fig. S3 C). Among these, TGF-β is a key inducer for epithelial-mesenchymal transition of cancer cells, thus confirming the results of Hallmark enrichment. AKT up and mTOR down indicates inhibition of the PI3K-AKT-mTOR pathway, which might lead to cell proliferation inhibition (Supplementary Fig. S3).

In addition, we further performed GO MFs, CCs, BPs, KEGG pathway and Reactome enrichment based on all high score targets (Score>0.95) predicted by miRWalks 3.0 to avoid missing any key regulatory information resulting from merely taking the intersection of predicted targets into consideration[30]. All 2735 target genes with miRWalks Score>0.95 were used as the input genes of the functional enrichment analysis. The results, compared with the 275 overlapping genes' enrichment results, were a little different. In MFs, terms like ubiquitin protein ligase activity and enzyme binding were significantly enriched (Supplementary Fig. S2 A), indicating a close

association between miR-654-5p and the ubiquitination process. In GO BPs, except for cell-cell adhesion and MAPK cascade, terms such as ubiquitin-dependent protein catabolic process and protein targeting were most significantly enriched (Supplementary Fig. S2 C). For GO CC, SNARE complex, miRNA-related RISC complex, RISC-loading complex and ciliary base, and endoplasmic reticulum-Golgi intermediate compartment membrane were enriched (Supplementary Fig. S2 B). In KEGG pathway enrichment, terms such as ubiquitin mediated proteolysis, neurotrophin signaling pathway and cancer-related items such as glioma, platinum drug resistance, hedgehog-signaling pathway and hepatocellular carcinoma were enriched (Supplementary Fig. S2 D), which indicated that miR-654-5p might be involved in various aspects of carcinogenesis and tumor progression.

2.3.2 *In vitro* validation of miR-654-5p functions

In previous functional enrichment analysis, multiple results indicated that the targets of miR-654-5p were strongly associated with the epithelial-mesenchymal transition (Supplementary Fig. S3A). By analyzing its detailed regulatory pathways in KEGG pathways, we found that miR-654-5p might be an inhibitor of EMT. Epithelial-mesenchymal transition can be characterized via expression alteration of EMT-related signatures, which included epithelial status markers such as E-cadherin, ZO-1 and mesenchymal status markers such as vimentin, N-cadherin and Snail[31-33]. To explore whether miR-654-5p can regulate EMT, we upregulated and downregulated miR-654-5p in A549 cells by agomir (chemically modified miRNA mimics) and antagomir (chemically modified miRNA antisense), respectively, and collected the cells for western blotting detection of EMT-related markers. The results showed that the evaluated miR-654-5p could downregulate the expression of N-cadherin (to 0.58-fold, Fig. 5 B) and Snail (to 0.25-fold, Fig. 5 B), while enhancing the expression of E-cadherin (by 1.28-fold, Fig. 5 B), ZO-1 (to 1.68-fold, 5 B) and Claudin-1 (to 4.85-fold, 5 B), indicating that overexpression of miR-654-5p inhibited the EMT process in lung adenocarcinoma cells. In miR-654-5p-knockdown cells, the results indicated the downregulation of epithelial markers (to 87% for E-cadherin, 95% for ZO-1, and 26% for Claudin-1; Fig. 5 B) and the upregulation of mesenchymal markers (by 406% for N-cadherin and 74% for Snail; Fig. 5 B).

In previous functional enrichment analysis, miR-654-5p was found to be involved in cell proliferation-related signaling pathways such as the VEGFR signaling pathway, RTK signaling pathway and cellular response to growth factor stimulus, etc. (Fig. 4 A, 4 D; Supplementary Fig. S2 A, S2 D , S3 D, S3 E). We therefore suggest miR-654-5p might be a crucial suppressor of cell proliferation, which were further verified by KEGG pathway mapping (Supplementary Fig. S4). In addition, enrichment analysis indicated that cell adhesion was a potential function of miR-654-5p, while the epithelial-mesenchymal transition (EMT), a pivotal process of cancer metastasis, is known to decrease cell-cell adhesion and give cancer cells migratory and invasive capacity[31,32,34,35]. We reasonably concluded that miR-654-5p might regulate cell migration and invasion via regulation of the EMT process.

To explore the effect of miR-654-5p on the migration and proliferation potential of lung cancer cells, we overexpressed miR-654-5p in A549 cells via the transfection of miR-654-5p agomir and stable N.C.

Then, 36 hours after transfection, CCK8 cell proliferation assays, wound healing assays were subsequently performed to assess the proliferative and migratory capacity of the cells. The results showed that the miR-654-5p-overexpression group showed higher proliferative ability (Fig. 5 C, $p < 0.01$) and mobility (Fig. 5 E, $p < 0.01$) compared with the control group transfected with stable N.C, indicating that upregulation of miR-654-5p inhibits the migration capacity of lung cancer cells. To further prove our hypothesis, H1299 cells were transfected with chemically modified antagomir to knockdown the endogenous abundance of miR-654-5p. Proliferation and migration assays were then performed as well, and the results showed that decreased miR-654-5p promoted the proliferation capacity (Fig. 5 D, $p < 0.01$) and migration capacity (Fig. 5 H, $p < 0.01$) of lung cancer cells. In summary, these results showed miR-654-5p inhibited lung cancer cell migration and proliferation *in vitro* and proved our functional enrichment analysis as proofs of concept.

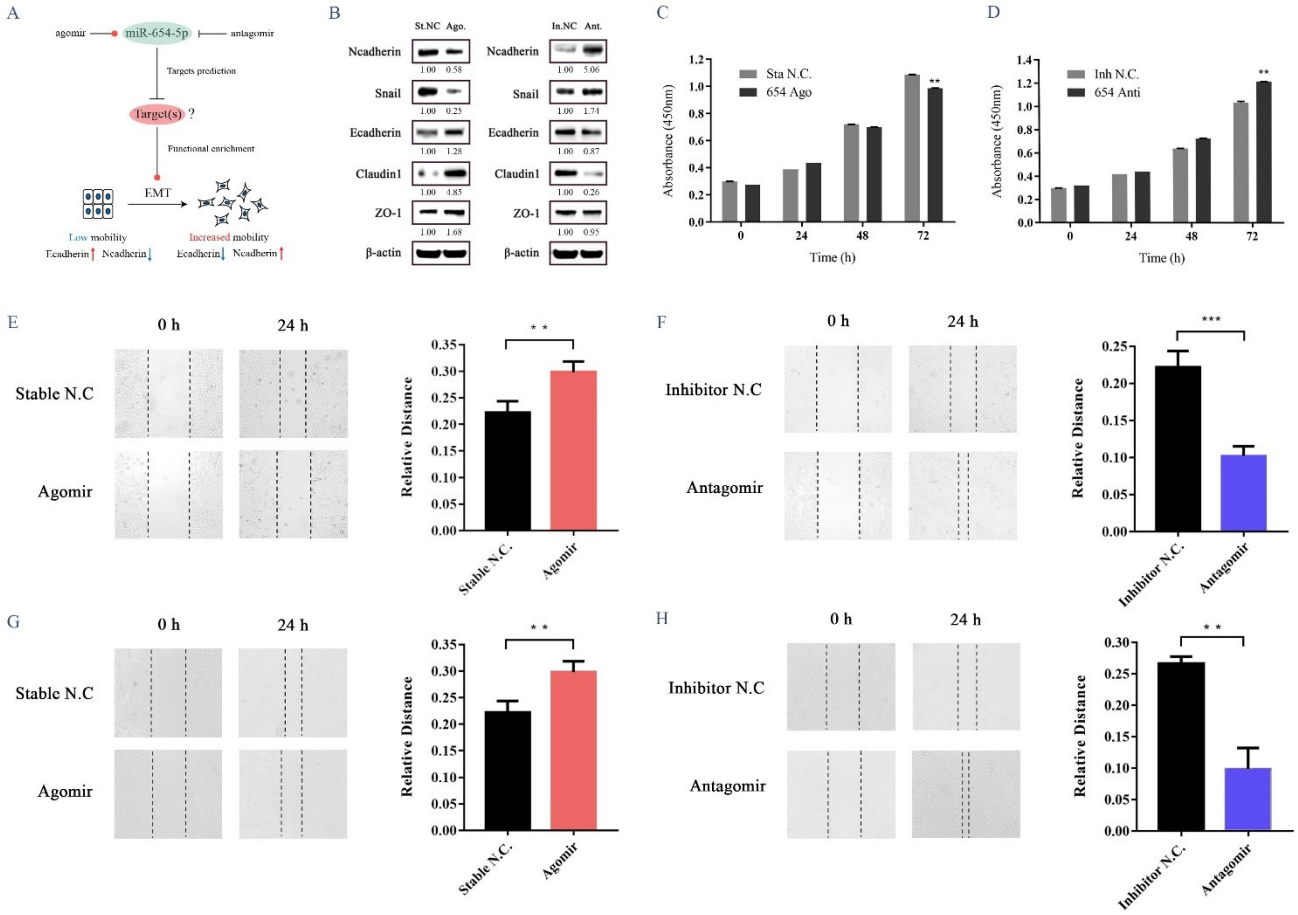

Figure 5. miR-654-5p inhibits lung cancer cell EMT, proliferation and migration. (A) Graphical representation of experimental validation of miR-654-5p functions *in vitro*. (B) A549 and H1299 cells were transfected with miR-654-5p agomir or antagomir respectively. 36 h after transfection, the cells were harvested and lysed to extract total proteins, then western blot assays were performed to detect changes in EMT-related markers (epithelial status hallmarks: E-cadherin, ZO-1 and claudin-1; mesenchymal status hallmarks: Ncadherin and Snail). (C and D) A CCK8 cell vitality assay was performed to detect the cell proliferation capacity of miR-654-5p-overexpressing A549 cells (C) and miR-654-5p-knockdown H1299 cells (D). (E and F) Representative microscopic images and quantitative results of migratory cells from the miR-654-5p-mimic (agomir)-transfected A549 lung cancer cell group (E) and the miR-654-5p-inhibitor (antagomir)-transfected A549 lung cancer cell group (F) in wound healing

assays; (F and G) Representative microscopic images and quantitative results of migratory cells from the miR-654-5p-mimic (agomir)-transfected H1299 lung cancer cell group (E) and the miR-654-5p-inhibitor (antagomir)-transfected H1299 lung cancer cell group (F) in a wound healing assays;. Data from the CCK8 proliferation assay and wound healing assay represent the mean±SEM of 3 independently prepared samples. *$p<0.05$, **$p<0.01$, ***$p<0.001$ vs. control.

2.4 Identification and clinical value of hub target genes

2.4.1 Construction of PPI network and identification of hub target genes

To further elucidate the potential interactions among the 275 overlapping genes, a protein-protein network was constructed utilizing the STRING database (Supplementary Fig. S5). With a high confidence, a network including 98 nodes and 120 edges was constructed and visualized (Fig. 6 A). These 98 genes might be crucial in miR-654-5p-regulated cellular processes especially those highly connected to others such as PIK3R1 and RHOCA. By combining these 98 genes with those differentially expressed in lung adenocarcinoma (LUAD) and those highly connected (Degree >2), we found that there are 11 genes in the intersection including HIST2H2BE, RGS4, and RAB10. This finding indicates that these predicted targets of miR-654-5p might play pivotal roles in lung adenocarcinoma (Fig. S6). These genes were selected as hub genes to perform further research.

To find the potential interconnected regions in this network, MCODE was utilized and five networks were clustered (Fig. 5 B-F). In these networks, AKT3, RAB1B, RTF1, EXOC7, and ALPL were identified as seeds, indicating these genes might be the promising pivotal genes in miR-654-5p-regulated biological processes, which could be further selected to perform experimental validation.

On the other hand, 275 targets of miR-654-5p with differential expression and differential survival rate in various cancers were also screen in order to provide more specific suggestions for miR-654-5p functional researches in various cancers (Supplementary Fig. S7 and Supplementary Table S1).

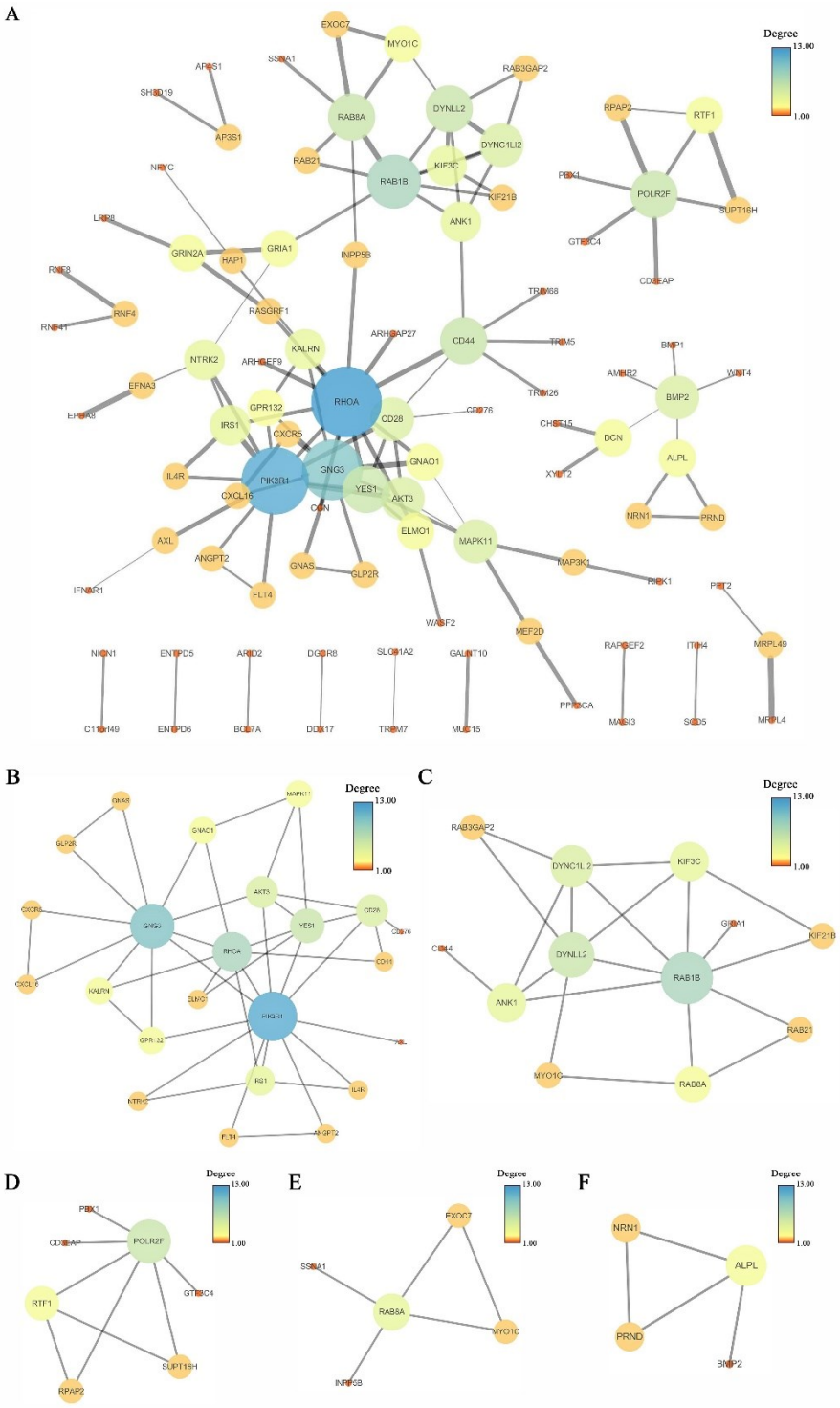

Figure 6. The PPI network of overlapping genes. (A) A network with 98 nodes and 120 edges was constructed by the STRING database for the overlapping 275 genes predicted by five promising miRNA-targets prediction tools with a high confidence (interaction score>0.700). Disconnected nodes were removed. Edges connecting nodes symbolized the interaction between two genes and the wider edges indicated a higher combined score between two nodes. The color and size of the nodes indicates the degree of the nodes. The network was visualized by Cytoscape 3.7.1 software. (B-F) built-in app MCODE was utilized to find clusters in the whole network. Five clusters were found.

2.4.2 Pan-cancer expression analysis of hub genes

To understand the roles of the 11 hub targets in various cancers, we analyzed the expression patterns of the hub genes in cancer and normal tissue in the TCGA database. The pan-cancer result is shown in Fig 7 B, which provides an expression matrix heatmap based on the given hub gene list. The heatmap showed that these hub genes exhibited distinct expression patterns in different cancer. For example, most hub genes were downregulated in LUAD, BLCA and BRCA (Fig 7 A and B), while most genes were upregulated in PAAD, indicating the roles of miR-654-5p/targets/axes were different in various cancers. Interestingly, a similar expression pattern of hub genes could also be found. For instance, the hub genes expression pattern in THYM and DLBC, showing the unification of correlations between miR-654-5p/targets/axes and a certain type of cancer.

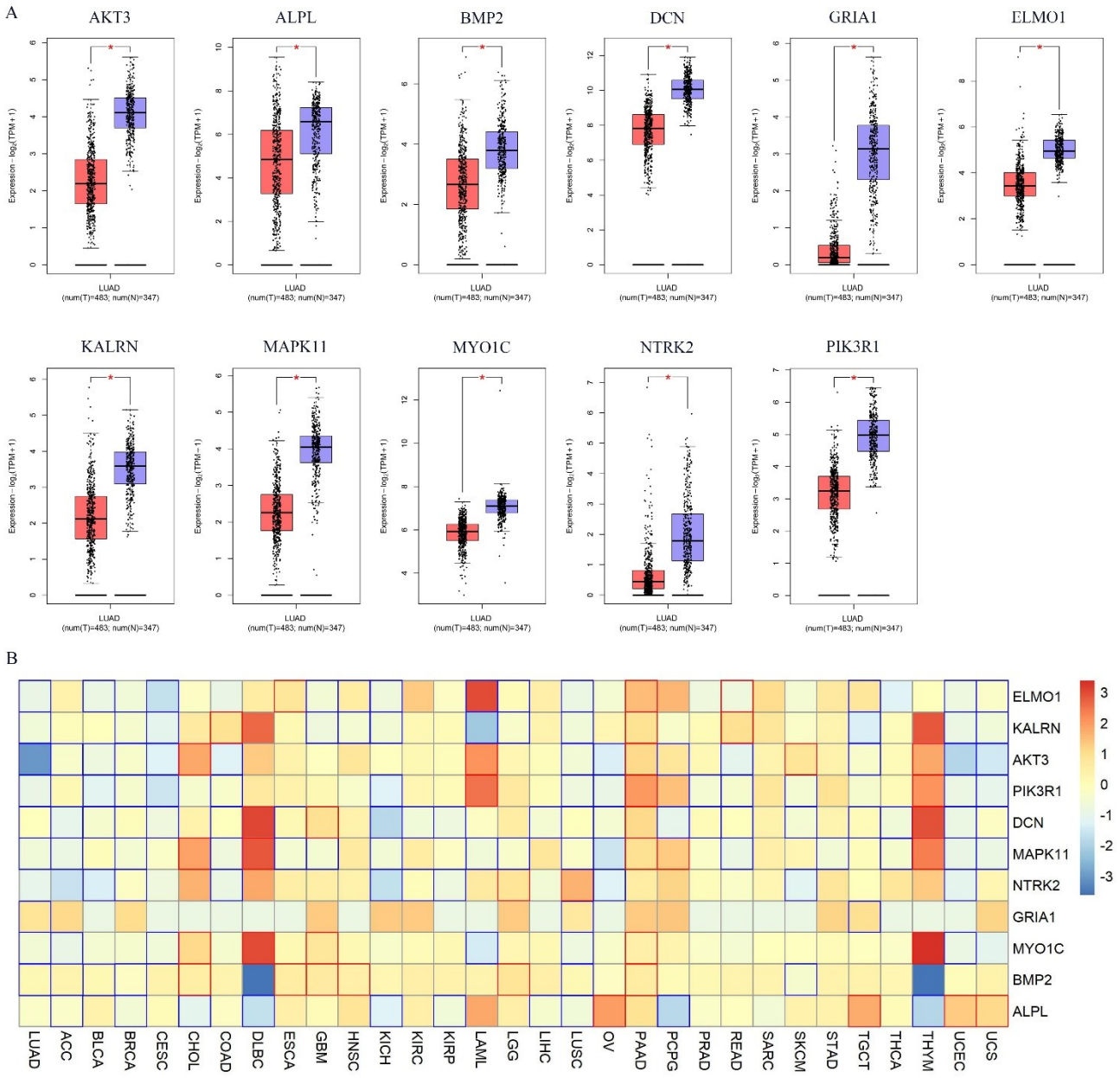

Figure 7. The pan-cancer differential expression analyses of 11 hub targets. (A) The differential expression analyses of hub genes in LUAD was based on TCGA samples (Tumor: 483 samples; Normal: 347 samples); their expression is presented in a Log2(TPM + 1) scale. (B) Pan-cancer differential expression heatmap was normalized and plotted by R package "pheatmap" based on TCGA cancer samples. The color in each block represents the median expression ratio of a gene (tumor/normal). Differentially expressed genes are framed in red (upregulated) or blue (downregulated), a P-value<0.05 was considered to indicate a statistically significant difference.

2.4.3 The prognostic potential of hub genes

To further explore the prognostic value of the hub genes, Kaplan-Meier survival analysis was performed based on the 11 hub genes selected by the PPI network in the LUAD TCGA database, in order to acquire a promising judgement as to whether these hub genes can be prognostic markers. During the analysis, Jetset probes were used as the proper probe standing for the promising expression of specific genes[20]. The results showed that in LUAD, high expression of MAPK11 and MYO1C (Fig. 8 A and C) was negatively related to patients' survival rate, while high expression of GRIA1, PIK3R1 and ELMO1 was correlated with a better prognosis (Fig. 7 B, D and E). These genes can be potential prognostic markers for lung cancer. To explore the potential prognostic value of the hub genes, by testing combinations of these hub genes as prognostic signatures, we found that combining GRIA1 and PIK3R1 as a signature could be of great prognostic value (Fig 7 F). Furthermore, the survival rate heatmap including OS (Fig. 7 G) and RFS (Supplementary Fig. S8) of these hub genes was also plotted based on the TCGA database, which might be used as a reference of prognostic markers.

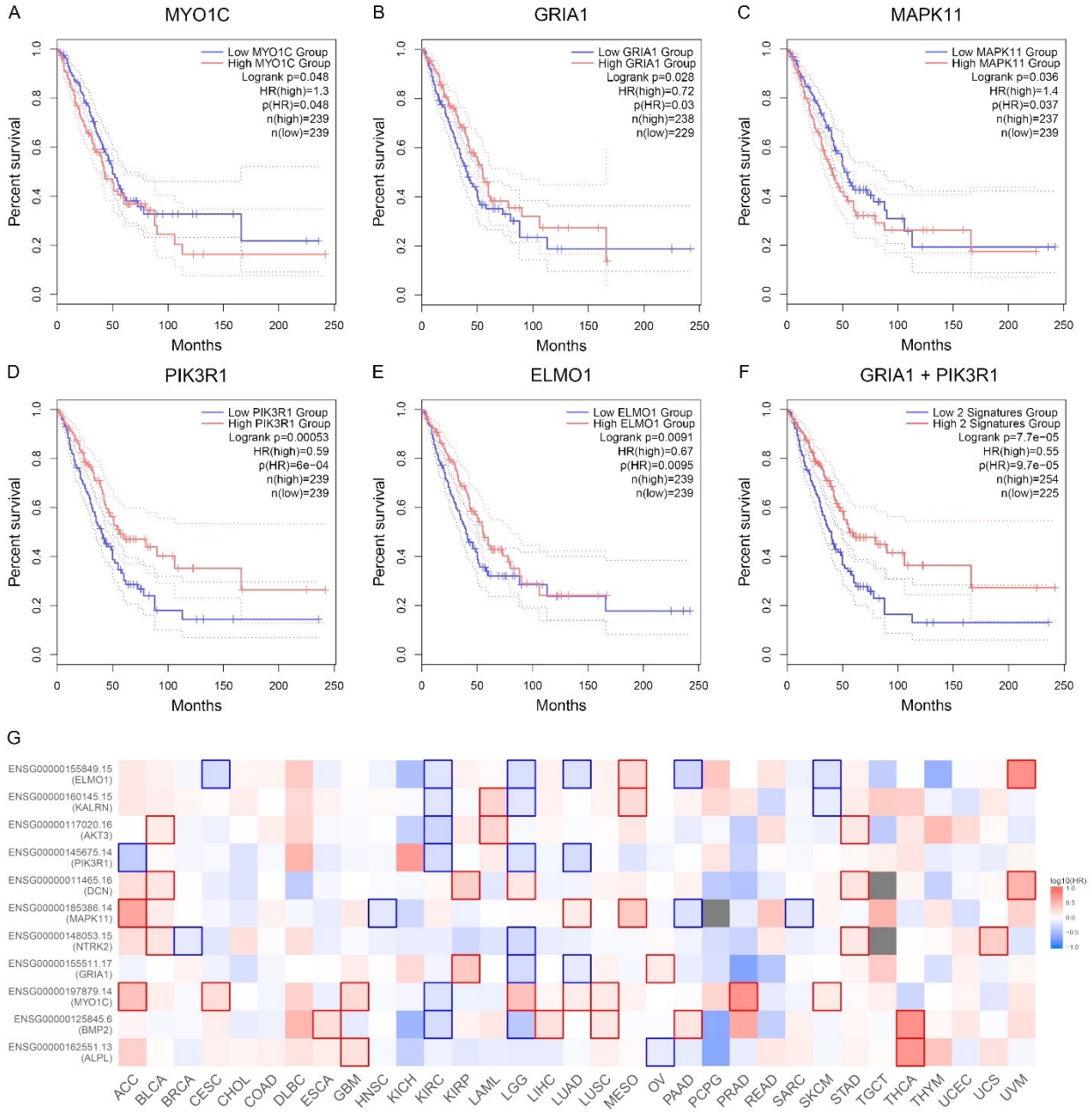

Figure 8. The overall survival rate of 11 hub targets in lung cancer and a pan-cancer survival heatmap. (A-F) The overall survival (OS) rate correlated with 11 hub targets by Kaplan-Meier survival analysis based on the TCGA LUAD database. A log rank p <0.05 was considered to indicate a statistically significant difference. Those with p>0.05 are not shown. (G) The heatmap of the pan-cancer OS rate of 11 hub targets by Kaplan-Meier survival analysis based on TCGA samples by GEPIA. A log rank p <0.05 was considered to indicate a statistically significant difference and are framed in red (positively correlated) or blue (negatively correlated).

3 Discussion

MiRNAs are a class of endogenous noncoding small RNAs that play an important role in post-transcriptional gene silencing via partially or completely complementary binding to target mRNAs in their 3' UTRs[23,36]. Currently, multiple researchers have confirmed that miRNAs are involved in the mechanisms of most biological processes including excessive growth, resistance to apoptosis, angiogenesis, invasion, and metastasis[35,37]. Accordingly, aberrant miRNA expression patterns are considered to be a sign of a variety of diseases such as cancer, suggesting the expression of miRNAs could be used as diagnostic, prognostic and predictive biomarkers[38]. The function of some miRNAs in cancer has been clarified and divided into oncogenic miRNAs (oncomiRs) or anti-oncogenic miRNAs (tumor suppressor miRs). For example, Let-7a, an identified tumor suppressor, has been shown to inhibit the proliferation of lung cancer cells by regulating HMGA2[39] and myc[40]. However, due to the complexity of mammalian cells, the miRNA regulatory network seems to be much more complex than a simple miRNA-target-phenotype regulation pattern. Therefore, identifying the target set of a miRNA and its potential functions would provide a larger landscape for understanding the mechanism underlying miRNA-regulated biological processes.

In our study, we focused on a poorly studied miRNA, miR-654-5p, and aimed to determine its main molecular mechanism and functions. Therefore, we designed a workflow integrating *in silico* identification, functional analysis, clinical value assessment of miR-654-5p hub targets and *in vitro* validation of targets and miR-654-5p functions (Fig. 1). First, utilizing four reliable miRNA target prediction tools including miRWalk 3.0[16], RNA22 2.0[41], miRanda[42,43] and TargetScan 7.2[44], we identified 275 overlapping targets that could be promising targets of miR-654-5p. To further assess this prediction method, these predicted targets were sorted by binding sites accessibility. As a result, RNF8, CYP4A11, ZNF385B, and WASF2 were ranked high in the gene list. We therefore chose RNF8 as an *in vitro* proof of concept for our predictions.

First, through a reverse target-miRNAs prediction using RNF8 3'UTR, we identified 25 potential miRNAs that regulated RNF8 and confirmed that miR-654-5p is one of them (Fig. 3 A and B) . Combined with the high accessibility of the RNF8 3'UTR binding sites, these results confirmed that the miR-654-5p-RNF8 axis is essential to the function of both miR-654-5p and RNF8. The main biological function of miRNA is to induce posttranslational repression or mRNA degradation of the target gene via base-pairing to miRNA recognition elements (MERs) located in the 3'-untranslated regions (3' UTRs) of target mRNAs[8]. We therefore screened the 3'UTR sequence of RNF8 mRNA, and two potential binding sites were found (Fig. 3 D). To prove the direct binding and translation repression effect of miR-654-5p on RNF8, 3980 nucleotides full-length RNF8 3'UTR were inserted into a dual-luciferase plasmid, and the results demonstrated miR-654-5p could regulate gene expression through direct binding to the RNF8 3'UTR. Then, we further proved that overexpression of miR-654-5p inhibited the protein expression level of RNF8, while knockdown of miR-654-5p by chemical-modified antisense (antagomir) enhanced the expression of RNF8. These results demonstrated the direct regulatory effect of miR-654-5p on RNF8 gene expression, which proved our target prediction strategy of miR-654-5p (Fig. 3).

To obtain a better and common understanding of regulatory pattern and functions of miR-654-5p at a cellular level, the previously predicted 275 overlapping target genes were subjected to GO annotation and KEGG pathway enrichment. We found that the MFs of these genes were more enriched in kinase activity like phosphotransferase activity and protein kinase activity. In addition, these 275 targets were mainly enriched by RTK-related biological processes such as cellular response to growth factor stimulus. In addition, growth factor-related and proliferation-related items were also enriched in the GO MFs (Fig. 4 A; Supplementary Fig. S2 A), GO CCs (Supplementary Fig. S3), GO BPs (Fig. 4 C; Supplementary Fig. S2 C), KEGG pathway (Fig. 4 D; Supplementary Fig. S2 D), Reactome (Supplementary Fig. S2 E and S3 D), and oncogenic signature (Supplementary Fig. S3 C), indicating miR-654-5p might be a crucial regulator of cell growth and proliferation. Therefore, we performed cell proliferation assays to assess the influence of miR-654-5p on cell proliferation *in vitro*. The results showed that miR-654-5p inhibited cell proliferative capacity, while downmodulation of miR-654-5p promoted cell proliferation.

In addition, we also noticed that items such as regulation of cell-cell adhesion were also enriched in GO BPs annotation (Fig. 4 C, Supplementary Fig. S2C), demonstrating miR-654-5p is likely to regulate cell adhesion . Consistently, in Hallmarks enrichment, 275 overlapping targets were enriched by epithelial-mesenchymal transition (EMT), which is a pivotal process of cancer metastasis known to decrease cell-cell adhesion and give cancer cells migratory and invasive capacity[31,32,34,35]. These results indicated that miR-654-5p might regulate cell adhesion via regulation of the EMT process. Interestingly, except for many enriched growth factor-related items, in the Reactome (Supplementary Fig. S3 D) and oncogenic signatures (Supplementary Fig. S3 C), targets were enriched in upregulated transforming growth factor-β (TGF-β, Supplementary Fig. S2 E and S3 C), a well-known inducer of EMT. Based on these findings, we conclude that miR-654-5p is likely to regulate the EMT process and therefore regulate cancer cell adhesion and migration. To further prove these findings *in vitro*, EMT-related hallmarks such as E-cadherin (also a hallmark of cell-cell adhesion), ZO-1, N-cadherin, and Snail were detected in miR-654-5p-overexpressing/knockdown lung cancer cells A549 and H1299 (Fig. 5). We found miR-654-5p indeed inhibited the EMT process and cell-cell adhesion. In addition, wound healing and Transwell assays proved the inhibitory function of miR-654-5p on cell migration, which confirmed our enrichment analyses results as an *in vitro* proof of concept.

Meanwhile, we also noticed that miR-654-5p might play important roles in many common biological processes. For example, neuron-related items such as dendrite and forebrain neuron development, dopaminergic synapse, and post-synaptic density were enriched (Fig. 4 B and C; Supplementary Fig. 2 C), indicating that miR-645-5p might be pivotal in neuron-related functions. Interestingly, a study focused on the correlation between miRNA and neurodevelopmental disorders revealed that miR-654-5p is commonly deregulated in autism spectrum disorders (ASD)[45], indicating an association between miR-654-5p and ASD, which needs further study.

In addition, bone development (Fig. 4 B and C), ubiquitin-related items such as ubiquitin protein ligase activity and ubiquitin-mediated proteolysis (Fig. 4 A; Supplementary Fig. 2 A, C and D), and cancers such as glioma and hepatocellular carcinoma were enriched. The relationship of miR-654-5p to these enriched items needs to be studied further. For pathways, we found MAPK-related signaling,

PI3K/AKT pathway (Supplementary Fig. 2 C, E; Supplementary Fig. 3 B, C, D, E and F), T cell receptor (TCR) signaling (Fig. 4 D; Supplementary Fig. 3 E), and Ras signaling (Fig. 4 D; Supplementary Fig. 2 A and D; Supplementary Fig. 3 C and E) were all enriched in multiple enrichment analysis, and these pathways could be given priority when we further study the regulatory pattern of miR-654-5p. Through literature searching, we found that some of our analyses have already been proven, such as the regulatory function of miR-654-5p on the MAPK pathway in OSCC[13] and osteogenic differentiation[46]. Thus, our functional analysis provides lots of correlations between miR-654-5p and common biological processes, and these *in silico* results need to be validated in further miR-654-5p-related research.

Next, we constructed a PPI network with 98 nodes to reveal a potential regulatory network among 275 overlapping genes. By the MCODE module in Cytoscape software, five key clusters were further identified, and AKT3, RAB1B, RTF1, EXOC7, and ALPL were identified as seeds in the clusters (Fig. 6), which needs to be further validated experimentally in miR-654-5p-regulated phenotype or gene functional research, because miR-654-5p could be a key regulator of these genes. By combining the differentially expressed genes and high degree, we identified 11 genes, which should be highly interconnected and associated with lung adenocarcinoma. Thus, pan-cancer expression and survival analyses were performed on these 11 genes to determine their correlation and prognostic value with various cancers. We found that most hub genes were downregulated in LUAD, BLCA and BRCA, (Fig 7 A and B), and upregulated in PAAD. Of note, the hub genes' expression patterns in THYM and DLBC were similar, which needs to be explored in the future. Meanwhile, we found that MAPK11, MYO1C, GRIA1, PIK3R1 and ELMO1 were significantly correlated with patients' survival and these 5 genes could be utilized as prognostic markers for LUAD. Additionally, the results showed that using GRIA1 and PIK3R1 together as a signature was of great value for the prognosis of LUAD. In fact, in addition to LUAD, these genes could be prognostic markers for many cancers (Fig. 8 A).

4 Conclusion

Our research utilized a close-loop experiment integrating bioinformatic analysis and *in vitro* experimental validation to explore the function of the poorly studied miR-654-5p, aiming to provide a larger landscape for the functions and regulation network of miR-654-5p. In addition, we, for the first time, experimentally validated the direct regulatory effect of miR-654-5p on RNF8 as proof of concept of our target prediction. We also validated the regulation of miR-645-5p on the lung cancer cell EMT process, cell proliferation and migration capacity for the first time, which not only proved our bioinformatic functional enrichments but also revealed multiple biological functions that miR-654-5p might regulate.

5 Conflict of Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

6 Author Contributions

CL, LM and LZ designed the experiments.CL performed bioinformatic analysis the experiments, miR-654-5p-RNF8 *in vitro* functional validation and wrote the manuscript. LM and JK performed and analyzed the *in vitro* data. CZ and JM contributed to the construction of dual-luciferase vector. XQ and XM contributed to functional analyses. CL, LM and LZ edited the manuscript. LZ contributed to funding acquisition. **All authors read and approved the final manuscript.**

7 Funding

This work was supported by a grant from the National Natural Science Foundation of China (No. 31870855), Natural Science Foundation of Hunan Province, China (No. 2020JJ5657) and the Natural Science Foundation of Hunan Province, China (No. 2019JJ50731)

8 Acknowledgments

None.

9 Ethics approval and consent to participate

Not applicable.

10 Consent for publication

All the authors have read and approved the paper and declare no potential conflicts of interest in the paper.

11 Abbreviations

MiRNAs, microRNAs; GO, Gene oncology; KEGG, Kyoto Encyclopedia of Genes and Genomes; TCGA, The cancer genome atlas project; GEO, Gene Expression Omnibus; LUAD, Lung adenocarcinoma; EMT, Epithelial-mesenchymal transition; 3' UTR, 3'-untranslated regions; STRING, the Search Tool for the Retrieval of Interacting Genes; PPI, Protein-protein interaction; GEPIA, Gene Expression Profiling Interactive Analysis; GTEx, Genotype-Tissue Expression; MFs, molecular functions; BPs, biological process; CCs, cellular compounds.

12 Availability of data and materials

All the data and materials supporting the conclusions are included in the main paper.

Reference

1 Shi, Z. M. *et al.* Downregulation of miR-218 contributes to epithelial–mesenchymal transition and tumor metastasis in lung cancer by targeting Slug/ZEB2 signaling. *Oncogene* **36** (2017).

2 Siegel, R., Naishadham, D. & Jemal, A. Cancer statistics, 2012. *CA: a cancer journal for clinicians* **62**, 10-29, doi:10.3322/caac.20138 (2012).

3 Chen, W. *et al.* Cancer statistics in China, 2015. *CA: a cancer journal for clinicians* **66**, 115-132, doi:10.3322/caac.21338 (2016).

4 Siegel, R. L., Miller, K. D. & Jemal, A. Cancer statistics, 2015. *CA: a cancer journal for clinicians* **65**, 5-29, doi:10.3322/caac.21254 (2015).

5 Molina, J. R., Yang, P., Cassivi, S. D., Schild, S. E. & Adjei, A. A. Non-small cell lung cancer: epidemiology, risk factors, treatment, and survivorship. *Mayo Clinic proceedings* **83**, 584-594, doi:10.4065/83.5.584 (2008).

6 Wang, J. *et al.* miR-455-5p promotes cell growth and invasion by targeting SOCO3 in non-small cell lung cancer. *Oncotarget* **8**, 114956-114965, doi:10.18632/oncotarget.22565 (2017).

7 Chatani, M. [The state of the art: radiation therapy for non-small cell lung cancer]. *Gan to kagaku ryoho. Cancer & chemotherapy* **24 Suppl 3**, 373-378 (1997).

8 Sun, W., Julie Li, Y. S., Huang, H. D., Shyy, J. Y. & Chien, S. microRNA: a master regulator of cellular processes for bioengineering systems. *Annual Review of Biomedical Engineering* **12**, 1 (2010).

9 He, Z. *et al.* Up-Regulation of MiR-452 Inhibits Metastasis of Non-Small Cell Lung Cancer by Regulating BMI1. *Cellular Physiology & Biochemistry International Journal of Experimental Cellular Physiology Biochemistry & Pharmacology* **37**, 387-398 (2015).

10 Bartel, D. P. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* **116**, 281-297 (2004).

11 Min, L., Liu, C., Kuang, J., Wu, X. & Zhu, L. miR-214 inhibits epithelial-mesenchymal transition of breast cancer cells via downregulation of RNF8. *Acta Bioch Bioph Sin* **51**, 791-798, doi:10.1093/abbs/gmz067 PMID - 31294443 (2019).

12 Tan, Y. Y., Xu, X. Y., Wang, J. F., Zhang, C. W. & Zhang, S. C. J. A. J. o. C. R. MiR-654-5p attenuates breast cancer progression by targeting EPSTI1. **6**, 522 (2015).

13 Lu, M. *et al.* miR-654-5p Targets GRAP to Promote Proliferation, Metastasis, and Chemoresistance of Oral Squamous Cell Carcinoma Through Ras/MAPK Signaling. **37**, dna.2017.4095 (2018).

14 Gyorffy, B. *et al.* An online survival analysis tool to rapidly assess the effect of 22,277 genes on breast cancer prognosis using microarray data of 1,809 patients. *Breast Cancer Res Treat* **123**, 725-731, doi:10.1007/s10549-009-0674-9 (2010).

15 Koch, A., De Meyer, T., Jeschke, J. & Van Criekinge, W. MEXPRESS: visualizing expression, DNA methylation and clinical TCGA data. *BMC Genomics* **16**, 636, doi:10.1186/s12864-015-1847-z (2015).

16 Sticht, C., De La Torre, C., Parveen, A. & Gretz, N. miRWalk: An online resource for prediction of microRNA binding sites. *PLoS One* **13**, e0206239, doi:10.1371/journal.pone.0206239 (2018).

17 Zhou, Y. *et al.* Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat Commun* **10**, 1523, doi:10.1038/s41467-019-09234-6 (2019).

18 Tang, Z., Kang, B., Li, C., Chen, T. & Zhang, Z. GEPIA2: an enhanced web server for large-scale expression profiling and interactive analysis. *Nucleic Acids Res*, doi:10.1093/nar/gkz430 (2019).

19 Xie, Z. C. *et al.* Investigation of miR-136-5p key target genes and pathways in lung squamous cell cancer based on TCGA database and bioinformatics analysis. *Pathol Res Pract* **214**, 644-654, doi:10.1016/j.prp.2018.03.028 (2018).

20 Li, Q., Birkbak, N. J., Gyorffy, B., Szallasi, Z. & Eklund, A. C. Jetset: selecting the optimal microarray probe set to represent a gene. *BMC Bioinformatics* **12**, 474, doi:10.1186/1471-2105-12-474 (2011).

21 Nagy, A., Lanczky, A., Menyhart, O. & Gyorffy, B. Validation of miRNA prognostic power in hepatocellular carcinoma using expression data of independent datasets. *Sci Rep* **8**, 9227, doi:10.1038/s41598-018-27521-y (2018).

22 Lowery, A. J., Miller, N., Mcneill, R. E. & Kerin, M. J. MicroRNAs as prognostic indicators and therapeutic targets: potential effect on breast cancer management. *Clinical Cancer Research* **14**, 360-365 (2008).

23 Bartel, D. P. MicroRNAs: Target Recognition and Regulatory Functions. *Cell* **136**, 215-233 (2009).

24 Chen, C., Xia, R., Chen, H. & He, Y. J. b. TBtools, a Toolkit for Biologists integrating various HTS-data handling tools with a user-friendly interface. 289660 (2018).

25 Conway, J. R., Lex, A. & Gehlenborg, N. UpSetR: an R package for the visualization of intersecting sets and their properties. *Bioinformatics* **33**, 2938-2940, doi:10.1093/bioinformatics/btx364 (2017).

26 Lee, H. *et al.* The DNA Damage Transducer RNF8 Facilitates Cancer Chemoresistance and Progression through Twist Activation. *Molecular Cell* **63**, 1021-1033 (2016).

27 Kuang, J. *et al.* RNF8 promotes epithelial-mesenchymal transition of breast cancer cells. *Journal of Experimental & Clinical Cancer Research* **35**, 88 (2016).

28 Lee, H., Ruan, D., He, J. & Chan, C. H. Two-faced activity of RNF8: What "twists" it from a genome guardian to a cancer facilitator? *Molecular and Cellular Oncology* **3** (2016).

29 Kuang, J. *et al.* RNF8 Promotes Epithelial–Mesenchymal Transition in Lung Cancer Cells via Stabilization of Slug. **18**, 1638-1649, doi:10.1158/1541-7786.MCR-19-1211 %J Molecular Cancer Research (2020).

30 Sticht, C., La Torre, C. D., Parveen, A. & Gretz, N. J. P. O. miRWalk: An online resource for prediction of microRNA binding sites. **13** (2018).

31 Kalluri, R. & Weinberg, R. A. The basics of epithelial-mesenchymal transition. *Journal of Clinical Investigation* **119**, 1420-1428 (2015).

32 Wang, Y. & Shang, Y. Epigenetic control of epithelial-to-mesenchymal transition and cancer metastasis. *Experimental Cell Research* **319**, 160-169 (2013).

33 Thiery, J. P., Acloque, H., Huang, R. Y. & Nieto, M. A. Epithelial-mesenchymal transitions in development and disease. *Cell* **139**, 871 (2009).

34 Chaffer, C. L. & Weinberg, R. A. A Perspective on Cancer Cell Metastasis. *Science* **331**, 1559-1564 (2011).

35 Hanahan, D. & Weinberg, R. A. Hallmarks of cancer: the next generation. *Cell* **144**, 646-674, doi:10.1016/j.cell.2011.02.013 (2011).

36 Liu, C. *et al.* Bioinformatic Identification of miR-622 Key Target Genes and Experimental Validation of the miR-622-RNF8 Axis in Breast Cancer. *Frontiers Oncol* **9**, 1114, doi:10.3389/fonc.2019.01114 PMID - 31709182 (2019).

37 Hausser, J. & Zavolan, M. J. N. R. G. Identification and consequences of miRNA–target interactions — beyond repression of gene expression. **15**, 599 (2014).

38 Iorio, M. V. & Medicine, C. M. C. J. E. M. MicroRNA dysregulation in cancer: diagnostics, monitoring and therapeutics. A comprehensive review. **4**, 143-159 (2012).

39 Chen, C.-C. *et al.* Aberrant let7a/HMGA2 signaling activity with unique clinical phenotype in\\r JAK2\\r -mutated myeloproliferative neoplasms. **102**, 509-518 (2017).

40 Kumar, M. S., Lu, J., Mercer, K. L., Golub, T. R. & Jacks, T. Impaired microRNA processing enhances cellular transformation and tumorigenesis. *Nat Genet* **39**, 673-677, doi:10.1038/ng2003 (2007).

41 Miranda, K. C. *et al.* A pattern-based method for the identification of MicroRNA binding sites and their corresponding heteroduplexes. *Cell* **126**, 1203-1217, doi:10.1016/j.cell.2006.07.031 (2006).

42 Betel, D., Koppal, A., Agius, P., Sander, C. & Leslie, C. Comprehensive modeling of microRNA targets predicts functional non-conserved and non-canonical sites. *Genome Biol* **11**, R90, doi:10.1186/gb-2010-11-8-r90 (2010).

43 Betel, D., Wilson, M., Gabow, A., Marks, D. S. & Sander, C. The microRNA.org resource: targets and expression. *Nucleic Acids Res* **36**, D149-153, doi:10.1093/nar/gkm995 (2008).

44 Agarwal, V., Bell, G. W., Nam, J. W. & Bartel, D. P. Predicting effective microRNA target sites in mammalian mRNAs. *Elife* **4**, doi:10.7554/eLife.05005 (2015).

45 Son, N. L. *et al.* Profiling olfactory stem cells from living patients identifies miRNAs relevant for autism pathophysiology. **7**, 1 (2016).

46 Wei, J. Q., Chen, H., Zheng, X. F., Zhang, B. X. & Li, T. S. J. J. o. S. M. U. [Hsa-miR-654-5p regulates osteogenic differentiation of human bone marrow mesenchymal stem cells by repressing bone morphogenetic protein 2]. **32**, 291-295 (2012).

47 Ai, C. & Kong, L. CGPS: A machine learning-based approach integrating multiple gene set analysis tools for better prioritization of biologically relevant pathways. *J Genet Genomics* **45**, 489-504, doi:10.1016/j.jgg.2018.08.002 (2018).

48 Xie, C. *et al.* KOBAS 2.0: a web server for annotation and identification of enriched pathways and diseases. *Nucleic Acids Res* **39**, W316-322, doi:10.1093/nar/gkr483 (2011).

Supplementary Information

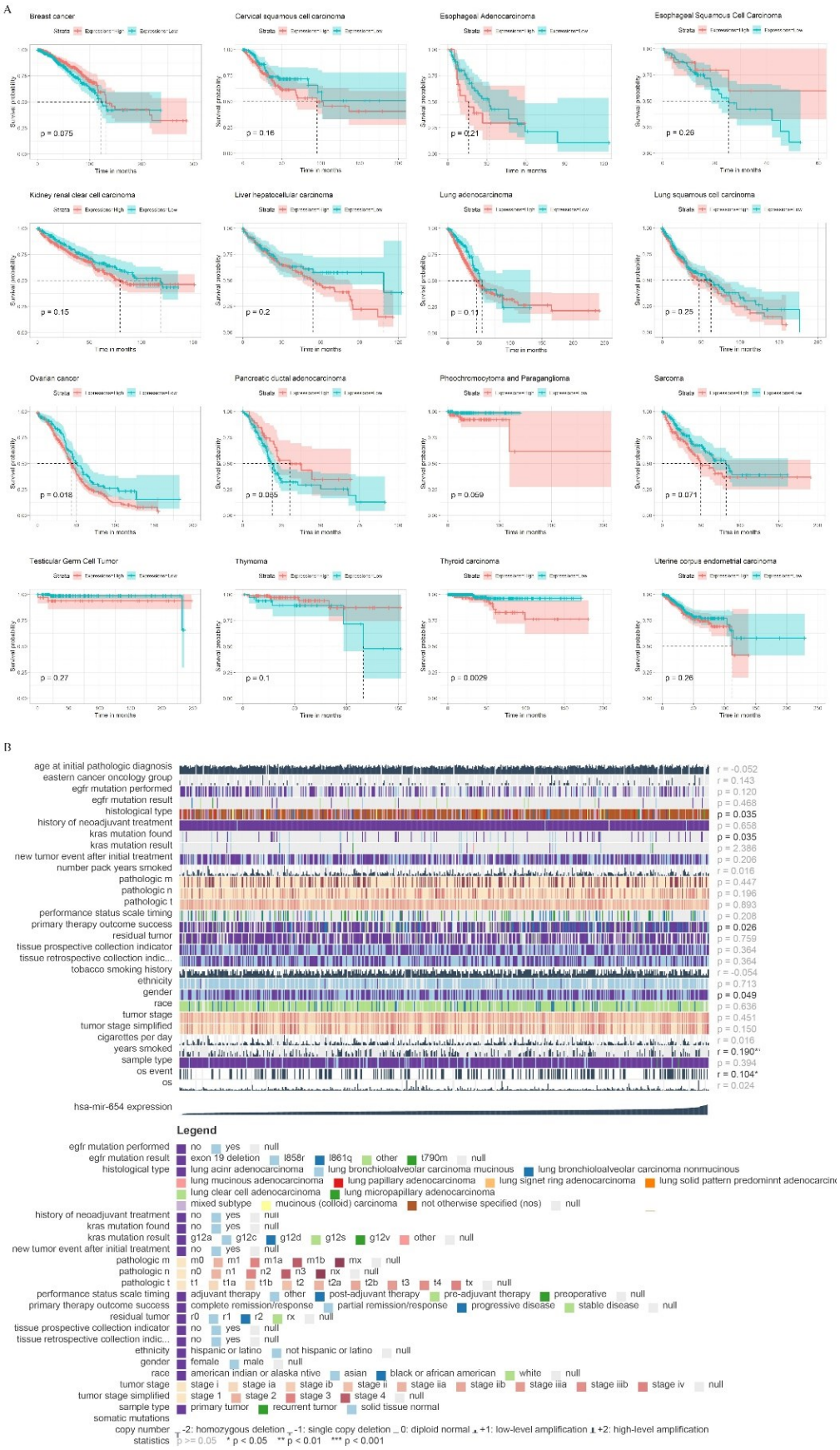

Figure S1. Survival analysis of hsa-miR-654. (A) Overall survival (OS) of hsa-miR-654 based on TCGA pan-cancer database[14]. (H-J) the correlation between miR-654 and clinical pathological features based on 861 TCGA LUAD samples[15].

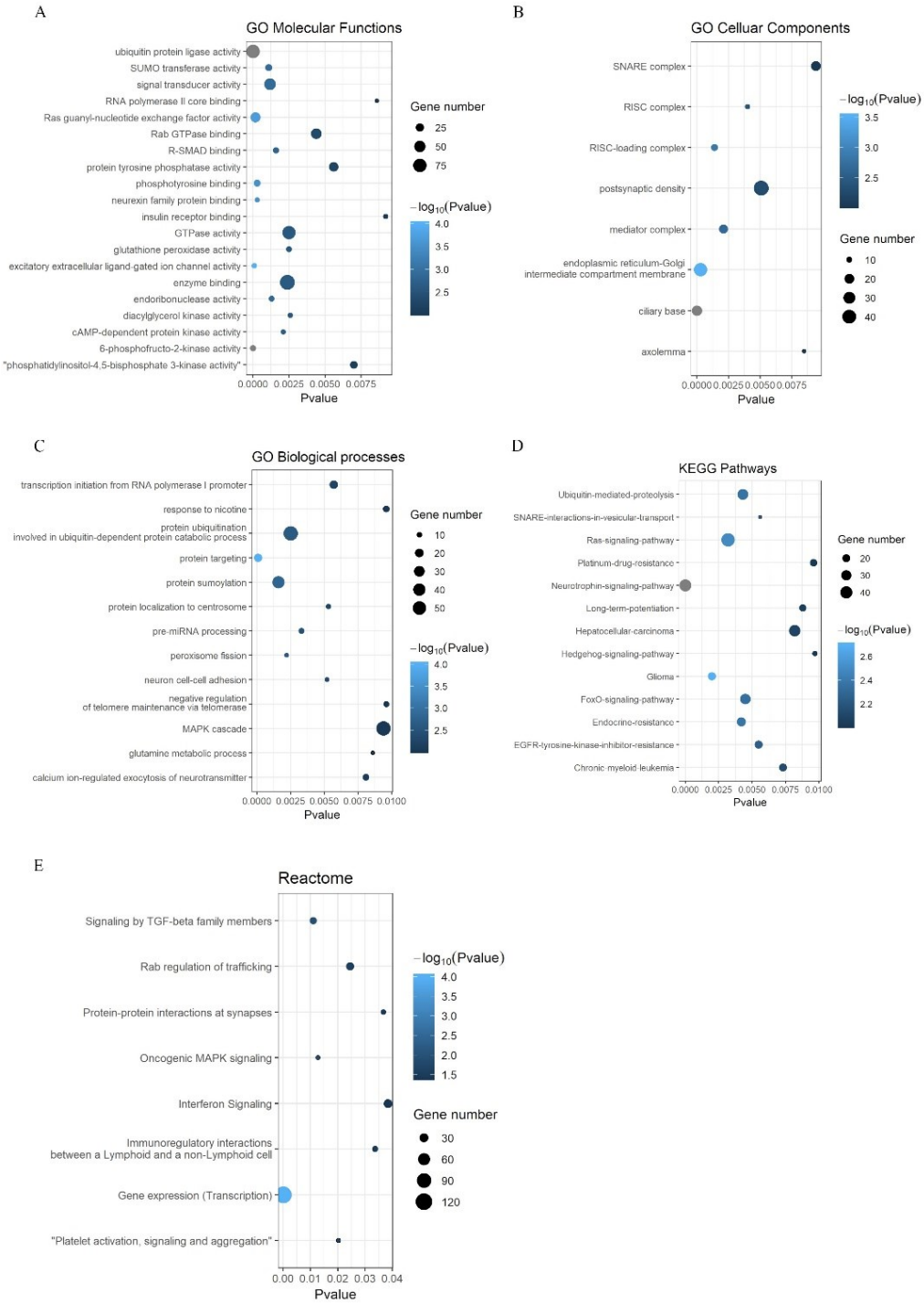

Figure S2. Gene ontology and KEGG pathway enrichment analysis of predicted targets of miR-654-5p by miRWalks3.0. the predicted genes (score>0.95) of miR-654-5p by new version 3.0 of miRWalks were all chosen to perform GO annotation and KEGG pathway enrichment analysis results to . Each bubble represents an enriched term, its size represents the counts of involved genes.

Lighter colors indicate smaller P values. (A) Enriched terms of GO molecular functions. (B) Enriched terms of GO cellular compounds. (C) Enriched terms of GO biological processes. (D) Enriched terms of KEGG pathway. (E) Enriched terms of Reactome.

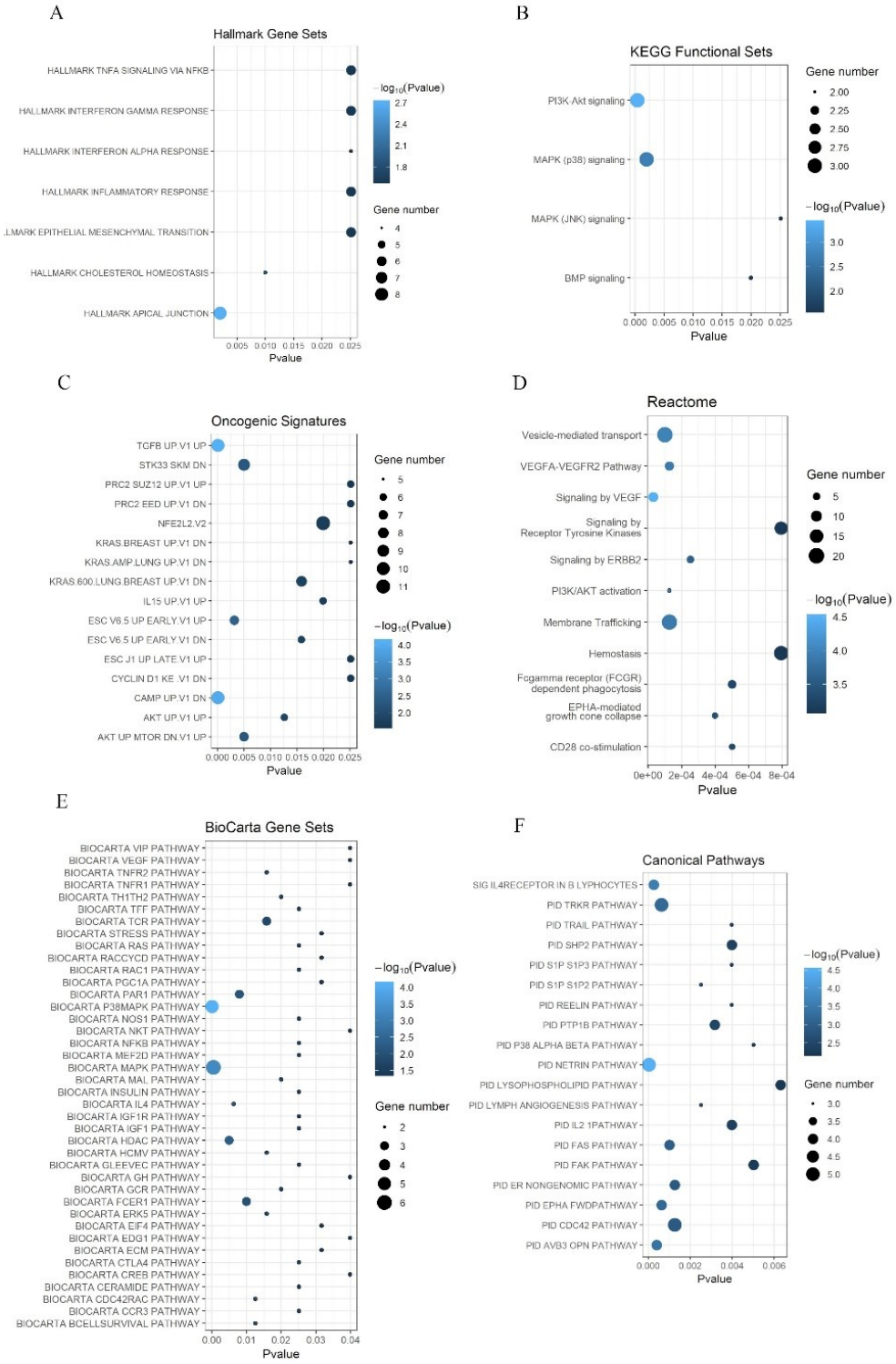

Figure S3. enrichment analysis of 275 overlapping predicted targets of miR-622. the 275 predicted genes of miR-654-5p were chosen to perform Hallmark, KEGG functional sets, Oncogenic signatures, Reactome, BioCarta and Canonical pathway enrichment analysis. Each bubble represents a term, and its size reprensent the counts of involved genes. Lighter colors indicate smaller P values. (A) enriched items for hallmark gene sets ($p<0.05$). (B) enriched items for KEGG functional sets ($p<0.05$). (C) enriched items for Oncogenic signatures ($p<0.05$). (D) enriched items for Reactome ($p<0.001$). (E) enriched items for BioCarta Gene Sets ($p<0.05$). (F) enriched items for Canonical Pathways ($p<0.01$)[47,48].

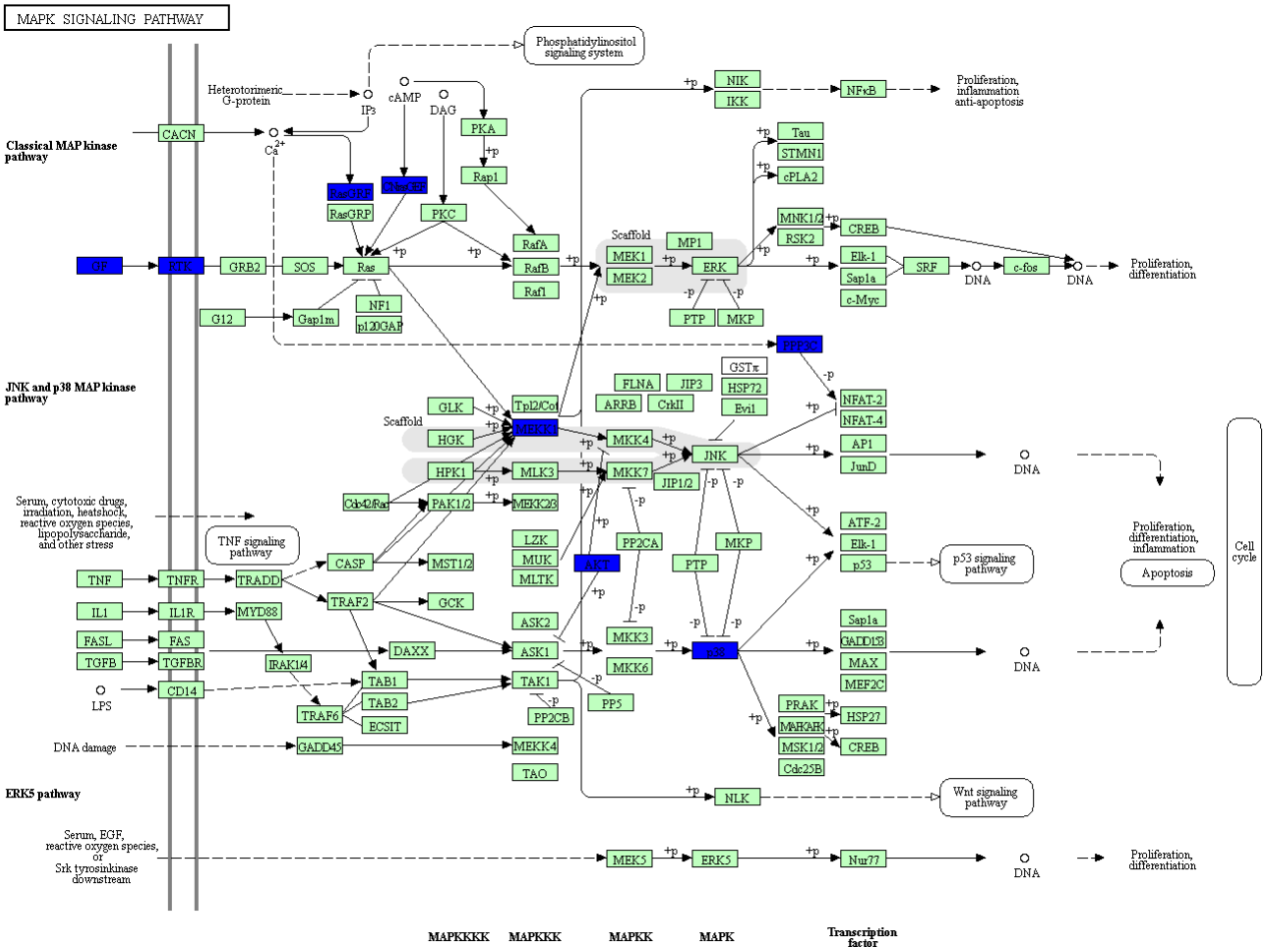

Figure S4. The inhibitory influence of miR-654-5p on MAKP pathway in *Homo sapiens*. the 275 predicted genes of miR-654-5p were mapped to KEGG pathway. Genes regulated by miR-654-5p were colored in blue.

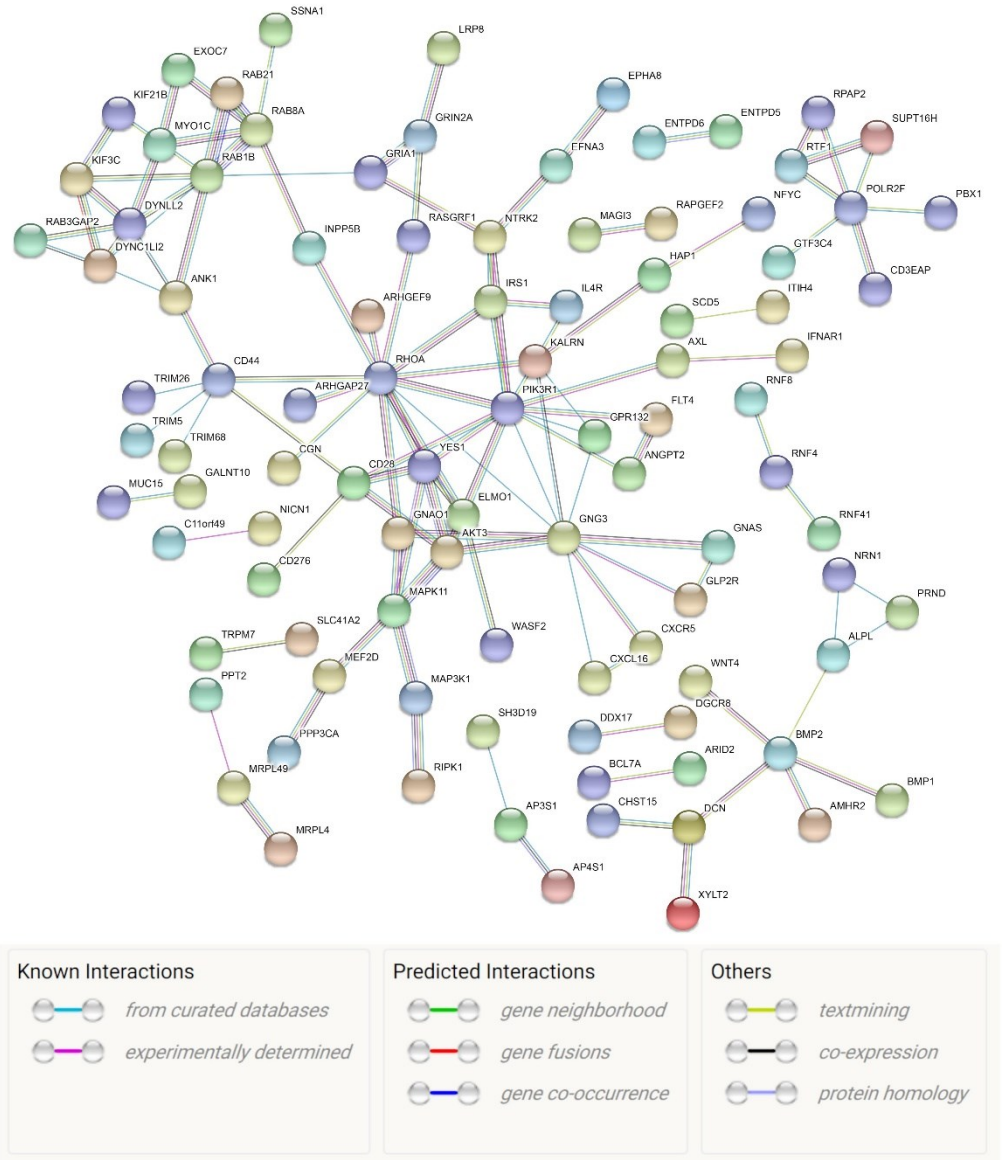

Figure S5. The PPI network of overlapping genes. A network with 98 nodes and 120 edges was constructed by the STRING database for the overlapping 275 genes predicted by five promising miRNA-targets prediction tools with a high confidence (interaction score>0.700). Disconnected nodes were removed.

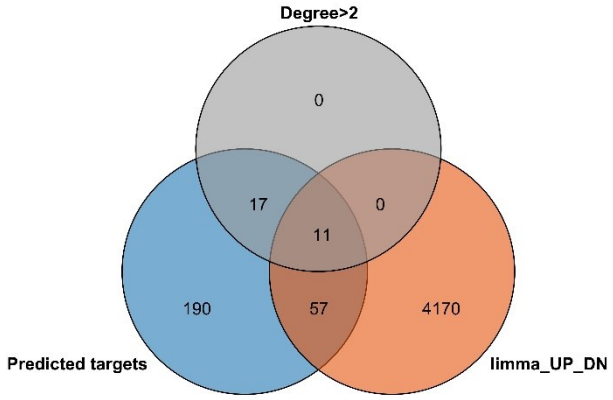

Figure S6. The hub genes of miR-654-5p in LUAD. 275 predicted genes were screened by the criterion of high connection in PPI network (Degree>2) and differential expressing (by R package limma with corrected p-value<0.05) in TCGA LUAD database[18].

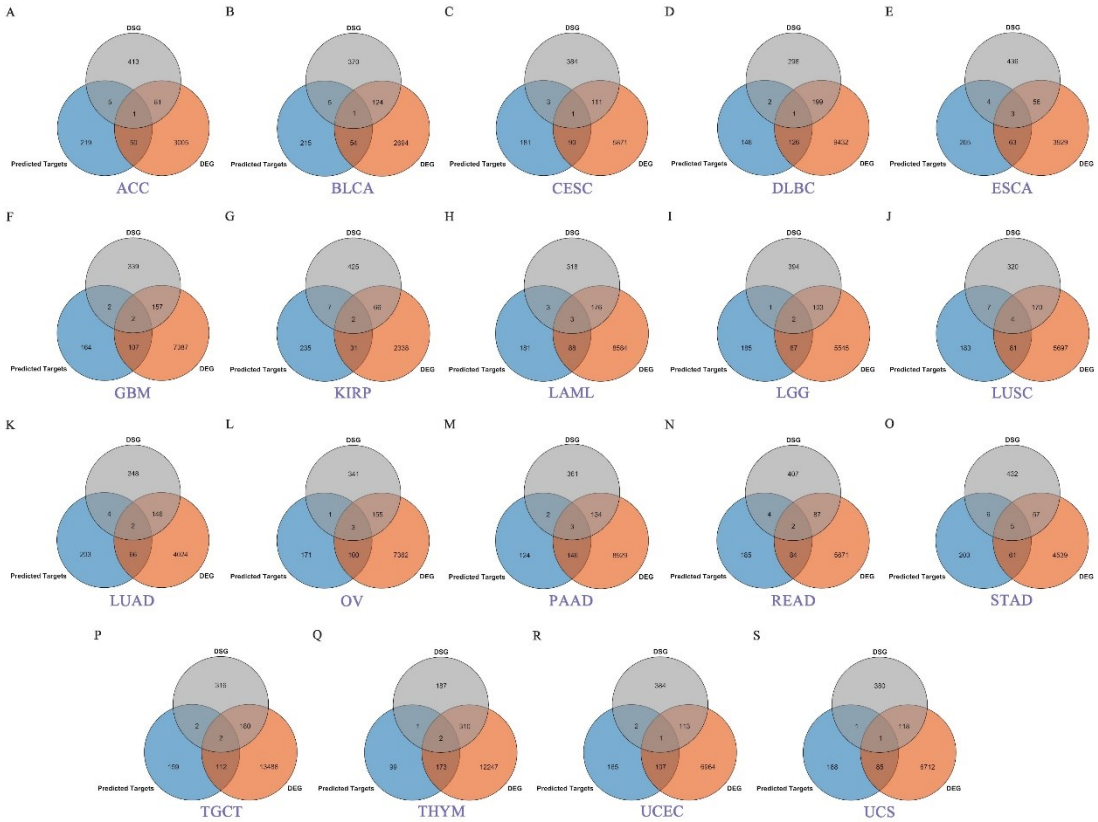

Figure S7. The identification of key targets of miR-654-5p in various cancers. 275 predicted genes were screened by the criterion of differential expressing (DEG, by R package limma with corrected p-value<0.05) and differential survival rate (DSG, ranked by OS p-value with median group cutoff, only top 500 were selected) in corresponding cancers based on TCGA databases[18].

ACC	BLCA	BRCA	CESC	COAD	DLBC	ESCA
ASF1B	ATP8B2	-	PEAR1	-	POLR2F	ATF3 GNL3L ANGPT2
GBM	**HNSC**	**KICH**	**KIRC**	**KIRP**	**LAML**	**LGG**
MYO1C TMEM150A	-	-	-	NRN1 MFAP4	RNF8 MRPL49 ASF1B	MAP3K1 PBX3
LIHC	**LUAD**	**LUSC**	**OV**	**PAAD**	**PRAD**	**READ**
-	PPT2 PIK3R1	CADM3 BMP2 FLT4 ASF1B	ALPL GALNT10 PEAR1	ATP8B2 AP3S1 ASF1B	-	ZNF493 NICN1
SKCM	**STAD**	**TGCT**	**THCA**	**THYM**	**UCEC**	**UCS**
-	EFNA3 HIF3A AFF3 UBE2QL1 ASF1B	PEAR1 ZNF385B	-	SLC7A6 HDAC7	PEAR1	SYT13

Table S1. Theidentification of key targets of miR-654-5p in various cancers.

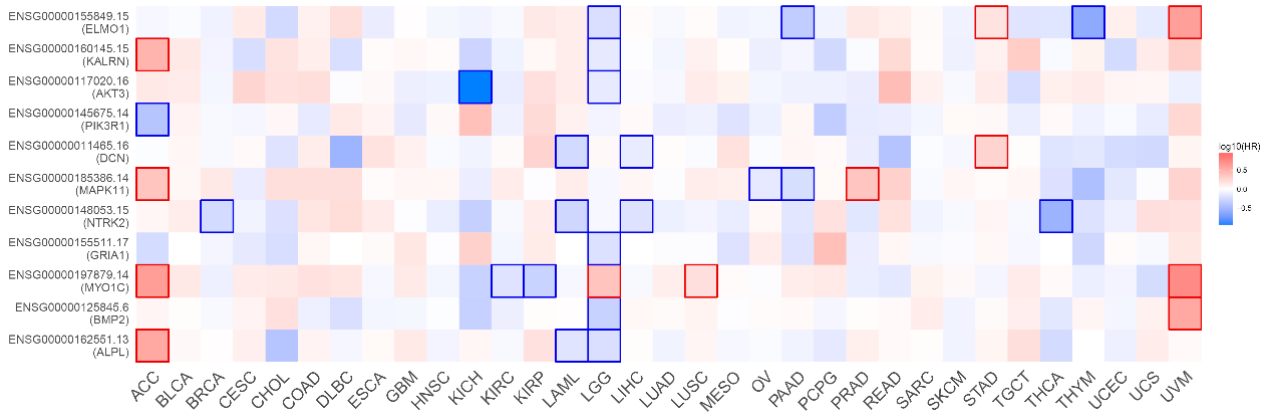

Figure S8. The relapse-free survival heatmap of 11 hub targets. The relapse-free survival (RFS) rate correlated with 11 hub targets by Kaplan-Meier survival analysis based on the TCGA pan-cancer database. A log rank p <0.05 was considered to indicate a statistically significant difference and are framed in red (positively correlated) or blue (negatively correlated).

Reference

1 Gyorffy, B. *et al.* An online survival analysis tool to rapidly assess the effect of 22,277 genes on breast cancer prognosis using microarray data of 1,809 patients. *Breast Cancer Res Treat* **123**, 725-731, doi:10.1007/s10549-009-0674-9 (2010).

2 Koch, A., De Meyer, T., Jeschke, J. & Van Criekinge, W. MEXPRESS: visualizing expression, DNA methylation and clinical TCGA data. *BMC Genomics* **16**, 636, doi:10.1186/s12864-015-1847-z (2015).

3 Ai, C. & Kong, L. CGPS: A machine learning-based approach integrating multiple gene set analysis tools for better prioritization of biologically relevant pathways. *J Genet Genomics* **45**, 489-504, doi:10.1016/j.jgg.2018.08.002 (2018).

4 Xie, C. *et al.* KOBAS 2.0: a web server for annotation and identification of enriched pathways and diseases. *Nucleic Acids Res* **39**, W316-322, doi:10.1093/nar/gkr483 (2011).

5 Tang, Z., Kang, B., Li, C., Chen, T. & Zhang, Z. GEPIA2: an enhanced web server for large-scale expression profiling and interactive analysis. *Nucleic Acids Res*, doi:10.1093/nar/gkz430 (2019).