scientific reports

Check for updates

OPEN

LINC00894, YEATS2-AS1, and SUGP2 genes as novel biomarkers for N0 status of lung adenocarcinoma

Marzyeh Alipour[1], Mehdi Moghanibashi[2]✉, Sirous Naeimi[3]✉ & Parisa Mohamadynejad[4]

Research on genes affecting tumors without lymph node metastasis is limited, hence this study employed both bioinformatic and experimental approaches to identify specific genes associated with lung cancer adenocarcinoma (LUAD) before lymph node metastasis occurs. Expression profiles of mRNAs and lncRNAs and LUAD clinical data were downloaded from the Cancer Genome Atlas (TCGA) using R software to identify differentially expressed genes (DEGs) associated with N0 and N + status. TargetScan, miRTarBase, and miRDB databases were used to identify interactions between miRNAs and mRNAs. The DIANA database and lncBase tool were used to find the association between lncRNA and miRNA. After selecting some genes, the expression of candidate genes was confirmed by real-time RT-PCR technique. Following the knockdown LINC00894 gene using the shRNA technique, its effect on invasion, migration, and apoptosis in the calu-3 cell line was investigated. In total, we found 321 specific DEGs not only in N0 vs. normal but also in N0 Vs. N + in LUAD, most of which were lncRNA and we identified a ceRNA network containing nine lncRNAs with the highest degree of connectivity. Among them, in addition to bioinformatic analyses, LINC00894 and YEATS2-AS1 were significantly increased in tumor tissues compared to normal tissues ($P = 0.0001$). also, SUGP2 that was shared in both lncRNA-related ceRNA subnetworks was significantly upregulated ($P = 0.0001$). Additionally, following the knockdown of LINC00894 in Calu-3 cell line, a significant decrease in migration and invasion was observed, but early apoptosis was significantly increased in the shLINC00894(48 h) group ($P = 0.007$). The findings of the present study show that lncRNAs play an important role in the N0 status of LUAD. Moreover, LINC00894, YEATS2-AS1, and SUGP2 can act as biomarkers in patients with N0 LUAD.

Keywords SUGP2, LINC00894, YEATS2-AS1, Lung adenocarcinoma, N0 lymph node metastasis, CeRNA network

Abbreviations

NSCLC	Non-small cell lung carcinoma
LUSC	Lung squamous carcinoma
LUAD	Lung adenocarcinoma
DEGs	Differentially expressed genes
TCGA	The Cancer Genome Atlas
LNM	Lymph Node Metastasis
AV/PI	Annexin V/ propidium iodide
R	Reverse
F	Forward
GAPDH	glyceraldehyde-3-phosphate dehydrogenase
lncRNA	Long non-coding RNA

[1]Department of Genetics, Colleague of Science, Kazerun Branch, Islamic Azad University, P.O. Code 7319866451, Kazerun, Iran. [2]Department of Genetics, Faculty of Basic Sciences, Kazerun Branch, Islamic Azad University, P.O. Code 7319866451, Kazerun, Iran. [3]Department of Biology, Zand Institute of Higher Education, Shiraz, Iran. [4]Department of Biology, Faculty of Basic Sciences, Shahrekord Branch, Islamic Azad University, Shahrekord, Iran. ✉email: mehdimoghani@yahoo.com; s.naiemi@zand.ac.ir

miR	microRNA
ROC	Receiver operating characteristics
qRT-PCR	Quantitative Real-time PCR
shCtrl	shRNA control
AUC	Area under the curve
MSigDB	Molecular signature database

Lung cancer has the highest mortality rate worldwide, and lung adenocarcinoma (LUAD) has become the predominant subtype of lung cancer[1]. LUAD is usually diagnosed at an advanced stage[2] and is characterized by aggressive progression, early metastatic spread, and high recurrence rates, causing significant morbidity and mortality across genders[3–5] and is genetically distinct from lung squamous cell carcinoma (LUSC)[6].

Studies have shown that patients with tumors appearing as N0 (without lymph node metastasis) on chest CT and PET/CT scans are typically characterized by their small size and minimal invasion of the surrounding area. Consequently, even if they contain hidden metastases, they can be surgically removed with a wedge or sublobar resection[7–9].

While various diagnostic methods, such as biopsy sampling, mediastinoscopy, and bronchoscopy, are available to identify LUAD, they are time-consuming and potentially delay immediate treatment initiation[10]. Also, low-dose CT scan, due to the low sensitivity of chest X-ray, has become the main way for early screening of lung cancer, but it is not accurate and can be harmful because growing radiation experience. In addition, histological analysis for tumor markers, such as CEA, is not only increased in lung cancer but can also be elevated in other tumors or diseases[11].

Therefore, identifying molecular biomarkers in the early stages of tumor development, before metastasis to lymph nodes and vessels, is an important prognostic factor for planning an effective treatment strategy[12].

Although previous studies have investigated genes related to lymph node metastasis (LNM) or metastasis in lung cancer[13,14], less attention has been devoted to genes specific to N0 status. Identifying potential tumor markers for N0 status is crucial to understanding the underlying molecular features of LUAD, which can contribute to the clinical diagnosis and treatment strategies for patients.

In addition to protein encoding genes (mRNAs), long non coding RNA (lncRNA) has been found to play a critical role in regulating various physiological and pathological cellular processes, including cancers[15]. One of the important roles of lncRNAs is to act as a sponge for miRNAs, binding and sequestering a significant number of them, which enables lncRNAs to compete with mRNAs for miRNA binding, and act as competitive endogenous or ceRNA and form a ceRNA network (ceRNA network) thereby influencing various biological activities[16].

In addition to experimental analyses, bioinformatics-based approaches provide a powerful method to discover novel genes and key pathways involved in cancer and to identify novel biomarkers for cancer diagnosis[17]. In this study, we first performed a comprehensive analysis of gene expression including mRNAs and lncRNAs in a large cohort of LUAD patients with and without lymph node metastasis using bioinformatics approaches to identify the most likely specific driver genes exclusive to N0 status, and then the candidate genes were validated by experimental methods.

Materials and methods

Data collection

A detailed overview of our study is presented in Fig. 1. The raw RNA sequencing data and the latest updated clinical data of the patients, including 59 normal, 278 N0 (LUAD patients without lymph node metastasis), and 122 N+ (LUAD patients with lymph node metastasis) samples of LUAD, were downloaded from the TCGA database (https://www.cancer.gov/ccg/research/genome-sequencing/tcga) using R software (https://r-pkgs.org/) and the TCGA biolinks package. Also, information regarding lncRNAs was extracted from the HUGO database (www.genenames.org) and used for the analysis. To ensure data consistency, the TCGA raw data were normalized using the trimmed mean of M-values (TMM) method, and the RNA expression levels were converted to log2 values.

Differential gene expression analysis in LUAD

With the limma package, we first screened differentially expressed mRNAs and lncRNAs for N0 and N + conditions of LUAD compared to normal samples using the linear model method, and then identified distinct mRNAs and lncRNAs that were exclusively associated with N0 status. Finally, we selected genes not only deregulated in N0 compared to normal samples but also significantly deregulated compared to N+. Statistical significance was defined as $P < 0.01$ and an absolute Log Fold Change ≥ 0. Given the complexity of the data and the large number of comparisons performed in our study, by setting a significance threshold of $P < 0.01$, rather than the conventional $P < 0.05$, we aim to minimize the likelihood of false positives and enhance the reliability of our findings. This is especially critical in high-throughput analyses or exploratory studies, where the likelihood of detecting spurious associations increases with the number of tests performed. This stringent criterion provides greater confidence in the biological relevance of the findings, particularly in the context of identifying biomarkers or potential therapeutic targets, as in our study. We visualized the differentially expressed RNAs through volcano plots using the ggplot software package.

Construction of the ceRNA network

We employed miRDB (http://mirdb.org/), miRWALK (http://mirwalk.umm.uni-heidelberg.de/), and LncBase V.3 (http://carolina.imis.athena-innovation.gr/diana_tools/web/) as online analytical tools to predict interactions

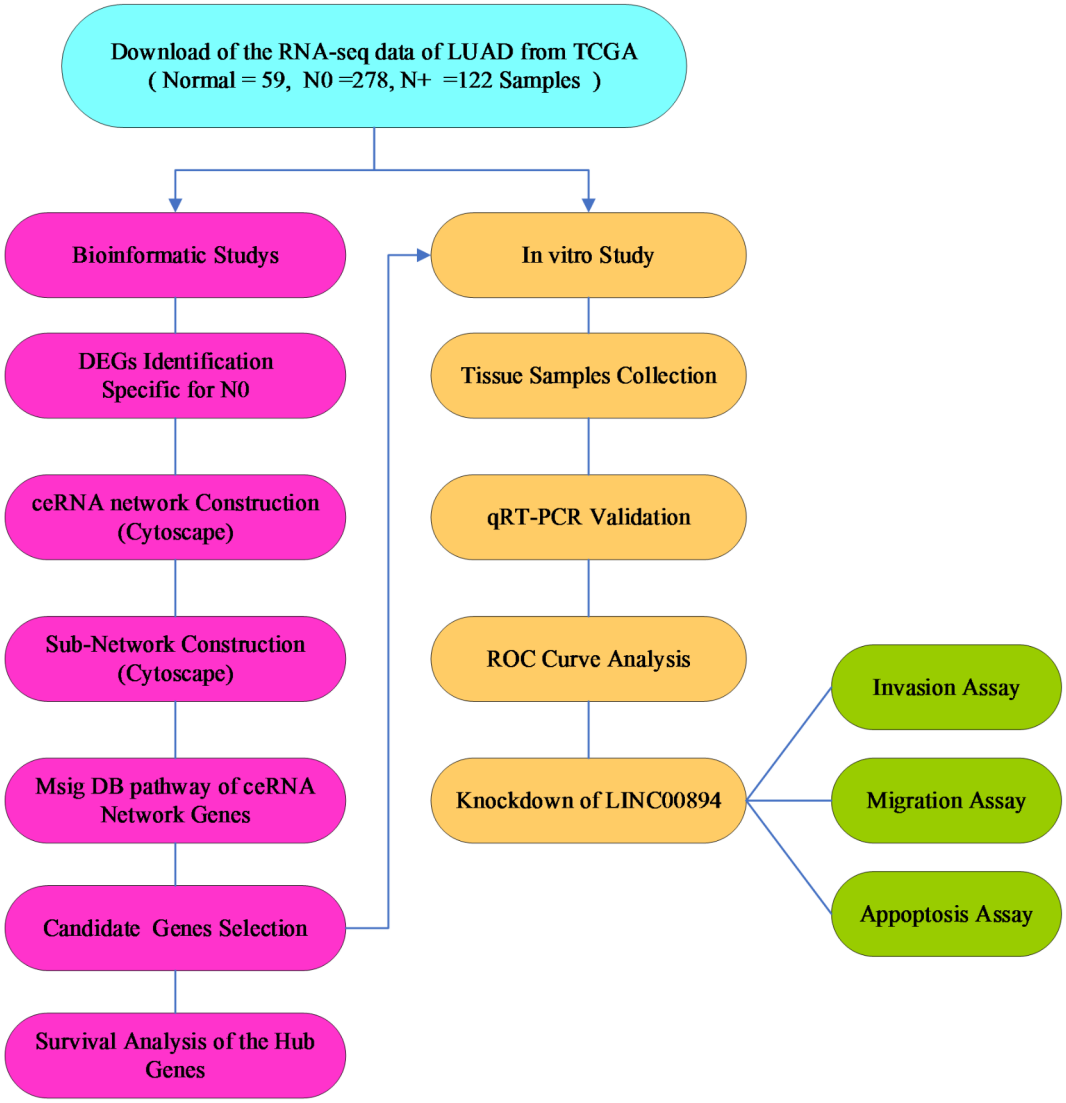

Fig. 1. The flow chart of the study including data collection, processing, and analyses.

between miRNAs and mRNAs, as well as lncRNAs and miRNAs. Next, we constructed a ceRNA network for N0-specific differentially expressed mRNAs and lncRNAs and visualized them using Cytoscape software (http://cytoscape.org/). Furthermore, we constructed and analyzed subnetworks associated with candidate lncRNAs using Cytoscape.

Survival analysis

To investigate the relationship between the expression of candidate genes and the prognosis of LUAD patients, we utilized the clinical TCGA data. The clinical data were subjected to preprocessing, which included removing normal samples. Subsequently, we extracted the expression of candidate genes for the samples that met the specified clinical conditions (selecting patients with lymph node metastasis or N + status and without lymph node metastasis or N0 status). To analyze the survival data, we employed the "survival" and "Survminer" packages in the R program (https://r-pkgs.org/). We generated Kaplan-Meier curves to visualize the survival outcomes.

Patients and tissue samples

Fresh tissue specimens were collected from patients diagnosed with LUAD, with ages ranging from 52 to 67 years and a mean age of 59 ± 2.2 years. The collection period spanned from 2020 to 2023 at hospitals in Esfahan, Iran, including Alzahra, Sina, Khanevadeh, and Milad. Inclusion criteria ensured that patients had not undergone radiotherapy, chemotherapy, immunotherapy, or other anticancer treatments before surgery. Tumor tissues, along with their corresponding adjacent tissues (≥ 1 cm), were obtained during resection and immediately stored in liquid nitrogen at −80 °C. Histological analyses confirmed the nature of the specimens, which were then categorized into three groups: N0, N+, and Normal (24 samples for each group). Informed consent was obtained

from all patients, who were fully informed about the study and willingly participated. The research protocol was approved by the Ethics Committee of Islamic Azad University, Kazerun branch (IR.IAU.KAU.REC.1400.142).

Cell culture

Two LUAD cell lines, A549 (N+) and Calu-3 (N0), as well as one normal human bronchial epithelial cell line (MRC-5), were obtained from the Cell Bank of the Danesh Bonyan Pasteur Institute of Iran. A549 cells were cultured in the RPMI-1640 medium, while Calu-3 cells and MRC-5 cells were cultured in the DMEM medium (BioIdea Company, Iran). The culture media were supplemented with 10% fetal bovine serum (FBS) (BioIdea Company, Iran) and 1% streptomycin-penicillin (BioIdea Company, Iran). The cells were maintained in an incubator at a constant temperature of 37 °C with 5% CO2. Before gene expression analysis, the cell lines were cultured for 6–8 passages.

RNA extraction and RT-qPCR assays

Total RNA was extracted from tissues and cells using the A101231 kit (PARSTOUS). The reverse transcription of mRNA and lncRNA was performed following the instructions provided with Easy cDNA reverse transcription Kit (A101161, PARSTOUS). For miRNA, the MicroRNA 1st Strand cDNA Synthesis Kit (Bon001027SL) was employed.

The Real-Time PCR System (Applied Biosystems™ (ABI), USA) was used for qRT-PCR assays to detect the levels of ANKRD10-IT1, YEATS2-AS1, LINC00894, hsa-let-7a-5p, hsa-let7e-5p, hsa-miR-423-5p, SUGP2, and MDM4 genes. Fast SYBR Green Master Mix was utilized as a reporter for detecting mRNA, miRNA, and lncRNA expression. The PCR reaction mixture underwent an initial incubation for 1 cycle at 95 °C for 15 min, followed by 40 cycles consisting of 95 °C for 20 s, 60 °C for 30 s, and 72 °C for 25 s. For the qRT-PCR, a reaction mixture for mRNA, lncRNA, and miRNA analysis containing 2 µl cDNA, 0.025 µl of each forward and reverse primer, 5 µl SYBR green Master Mix High ROX, and 4.5 µl of RNase-free water was prepared. The reaction mixture, with a final volume of 10 µl, was prepared on ice and subjected to specific time-temperature conditions. The forward and reverse primer sequences for GAPDH and candidate genes were designed using Oligo6 software, verified for integrity through NCBI BLAST, and synthesized by Macrogen Inc. The primer sequences used for qRT-PCR can be found in Table 1. We used U6 and GAPDH genes as Internal reference genes for the normalization of miRNAs and lncRNAs/mRNAs, respectively.

Knockdown of LINC00894 in Calu-3 cell line

Calu-3 cells were transfected with specific shRNAs targeting LINC00894 (shLINC00894) using the FavorPrep™ Plasmid DNA Extraction Mini Kit (Favorgen, Taiwan, cat no: FAPCK 000). Control groups included a non-targeting negative control (sh-NC) and an empty-vector control. Transfection of each plasmid into the cells was performed using Polyethyleneimine (PEI) according to the manufacturer's instructions. The transfection efficiency was evaluated after 24–48 h through fluorescence microscopy examination and quantitative reverse transcription-PCR (qRT-PCR).

shRNA transfection

The sh-LINC00894 and shNC sequences were designed using online shRNA tools available on biosettia.com (https://biosettia.com/support/shrna-designer). The specific sequences for the sh-LINC00894 and shNC transfections are as follows:

Sh lncRNA:

5′- GATCC CAGGCTGTGCAATCCATATTATTGGATCCAATAATATGGATTGCACAGCCTGA-3′

5′-AGCTTCAGGCTGTGCAATCCATATTATTGGATCCAATAATATGGATTGCACAGCCTGG-3′

shNC:5′- AATTCTCCGAACGTGTCACGT-3′

The short hairpin (sh)RNA and negative control (NC), were synthesized by the Pasteur Institute in Iran. The target sequences were incorporated into the PRNAT-H1-Neo vector through the respective restriction sites at both ends. Subsequently, the constructed vectors were transformed into TOP 10 E. coli competent cells, provided by the Pasteur Institute in Iran. Positive recombinants were identified through PCR screening. Plasmid DNA extraction was performed using the FavorPrep™ Plasmid DNA Extraction Mini kit (Favorgen Headquarters), and the concentration was determined using a spectrophotometer (Thermo Scientific™ NanoDropTM). Calu-3

Genes	Primer sequences (5′–3′)	Amplified fragment	Genes	Primer sequences (5′–3′)
YEATS2-AS1	F: GCCGTTTGTTCGTATCGCTG R: TTCCGTGTTCCTTTCCCGTG	137 bp	let7e-5p	F: GAT GGG CTG TGA GGTA R: Universal
LINC00894	F: TACACCGCCAAATCGCTAGG R: CGGGACAGGATCCGGTAAAC	118 bp	hsa-miR-423-5p	F: AGG GGC AGA GAG CG R: Universal
ANKRD10-IT1	F: GCACAGTCATCGTAGTTTCC R: AAACTCAGACAACGGTCAACT	151 bp	hsa-let-7a-5p	F: GGCTGAGGTAGTAGGTTGTAT R: Universal
SUGP2	F: TGCCCATCTGCTGACATTGA R: CCACAGGTCAGGGTTATCGG	129 bp	RNU6	F: AAG GAT GAC ACG CAA R: Universal
MDM4	F: AGACTGCCAAATTGCTCCCA R: ACGTTGCAGCAGAAACAAGC	190 bp	GAPDH	F: AAGCTCATTTCCTGGTATG R: CTTCCTCTTGTGCTCTTG

Table 1. Primers sequences of candidate genes for real-time RT-PCR analyses. R, reverse; F, forward; GAPDH, glyceraldehyde-3-phosphate dehydrogenase.

cells (3×10^5) were seeded in 6-well plates and transfected with the aforementioned plasmids or oligonucleotides using Polyethyleneimine reagent (PanReac Applichem), following the manufacturer's protocols. The time interval between transfection and subsequent experimentation was 48 h.

Assessment of cell migration, apoptosis, and invasion

Cell migration, apoptosis, and invasion were evaluated using the transwell migration assay, Annexin V/propidium iodide (AV/PI) staining, and Matrigel invasion assay, respectively. Transfected cells were seeded in the upper chambers of transwell inserts with serum-free medium, while the lower chambers were filled with medium containing 10% FBS. After a 24-hour incubation period, the cells that had migrated to the bottom surface of the membrane were fixed with 5% glutaraldehyde for 10 min and stained with 0.1% crystal violet for 20 min. The cells on the upper side of the membrane were carefully removed, and the migrated cells were counted in five random microscopic fields. For apoptosis assessment, Calu3 cells were plated in a six-well plate at a density of 5×10^5 cells and allowed to grow for 24 and 48 h. Harvested cells were washed with ice-cold PBS, resuspended in 100 µl of binding buffer, and stained with Annexin V/PI for 20 min in the dark under cold conditions. Flow cytometry (BD FACS Calibur) was used to analyze the apoptotic cells, and the data were analyzed using FlowJo v10.5.3 software. Furthermore, the invasion was evaluated using the Matrigel invasion assay. The cell invasive ability was assessed using Matrigel basement membrane matrix (Sigma Aldrich), Transwell filter chamber (FPL), glutaraldehyde solution (Merk, Germany), and crystal violet (Sigma Aldrich).

Statistical analysis

All pre-processing and statistical analysis, including the generation and visualization of graphs, were conducted using the R (version 4.2.2) and GraphPad Prism 5 software (version 8). ROC curves were plotted to assess the diagnostic performance of lncRNAs and mRNA, and a logistic regression model combined with ROC curve analysis was constructed using SPSS software and a nonparametric t-test. Statistical significance was defined as $P < 0.01$. The fold change expression of target genes in experimental analyses was determined using the $2^{-\Delta\Delta CT}$ method.

Results

Identification of N0-specific DEGs

Volcano plots of the differentially expressed lncRNAs and mRNAs for LUAD patients from TCGA for N0 status vs. normal samples and N + status vs. normal samples are shown in Fig. 2A-B. Using Venn diagrams, we further explored the similarities and specificities of the N0 and N + specific DEGs. The results showed that 2082 genes were specifically upregulated in the N0 status compared to normal (Fig. 2-C), while 706 genes were downregulated in the N0 status compared to normal (Fig. 2-D). Next, we selected genes that were not only deregulated specifically in N0 status but also up or down-regulated significantly compared to N+. Our analyses revealed that 278 genes were upregulated, while only 43 genes were downregulated (Fig. 2E-G) they are considered candidate genes for further analyses. Among them, 122 and 91 genes were classified as lncRNAs and mRNA, respectively.

Identification of several lncRNA–miRNA–mRNA ceRNA subnetworks specific to N0 LUAD

Among the 278 up and 43 downregulated genes specific to the N0 status of LUAD, 122 genes identified as lncRNAs and 91 genes as mRNAs (some of the N0-specific DEGs identified as pseudogenes or don't annotate). Considering that most of the identified DEGs were lncRNAs, and one of the significant functions of lncRNAs is their involvement in ceRNA networks, we constructed a ceRNA network for all N0-specific DEGs. We identified 427 and 438 miRNAs that interact with 122 N0-specific lncRNAs and 91 mRNAs, respectively (Fig. 3A). Among the lncRNAs, KCNQ1OT1, ANKRD10-IT1, LINC00342, THUMPD3-AS1, YEATS2-AS1, NEAT1, GHRLOS, LINC00894, and LINC01089 exhibited the highest degrees in the network (Table 2).

To gain further insight into the constructed network, we analyzed the signaling pathways of all genes within the ceRNA network. The results revealed that PI3K/AKT/mTOR pathway, cell division, and inflammation have the highest score, respectively (Fig. 3B).

Sub-network analyses among the highest-degree lncRNAs in the network showed that ANKRD10-IT1 (Fig. 4A), YEATS2-AS1 (Fig. 4B), and LINC00894 (Fig. 4C) have the most interaction with miRNAs and mRNAs. In addition, the review literature has shown that there is limited study on the role of these lncRNAs in LUAD, so we selected them together with SUGP2, MDM4, miR-423-5p, hsa-let-7a-5p, and hsa-let7e-5p, all of which were common to all three sub-networks, for further analyses.

Verification of the expression level of selected genes in LUAD tissues samples

To confirm our bioinformatics data for selected genes, we measured the expression levels of eight genes in LUAD tissues including 24 patients with N0 and adjacent normal tissues as well as 24 patients with N + status and adjacent normal tissues using real-time RT-PCR.

The expression levels of LINC00894 (Fig. 5A) and YEATS2-AS1 (Fig. 5B) were significantly upregulated compared to normal tissues ($p = 0.0001$), which is consistent with the bioinformatic analysis. However, the expression level of ANKRD10-IT1 (Fig. 5C) was not specific to N0 status, which contradicts the bioinformatic analysis. Also, there is no difference in the expression of hsa-let-7a-5p ($p = 0.8$) and miR-423-5p ($p = 0.2$) (Fig. 5D and E) in the tissues with N0 status compared to normal tissues. The expression level of hsa-let7e-5p (Fig. 5F) decreased in tissues with N0 status compared to normal tissues ($p = 0.03$) but it was not specific to N0 status. Furthermore, the expression level of SUGP2 (Fig. 5G) in N0 status showed a significant increase ($p = 0.0001$), which is consistent with the bioinformatic analysis and indicates its specificity to N0 status. Contrary to the bioinformatic prediction, there is no significant deregulation of the MDM4 gene ($p = 0.1$) (Fig. 5H) in the N0 status of LUAD.

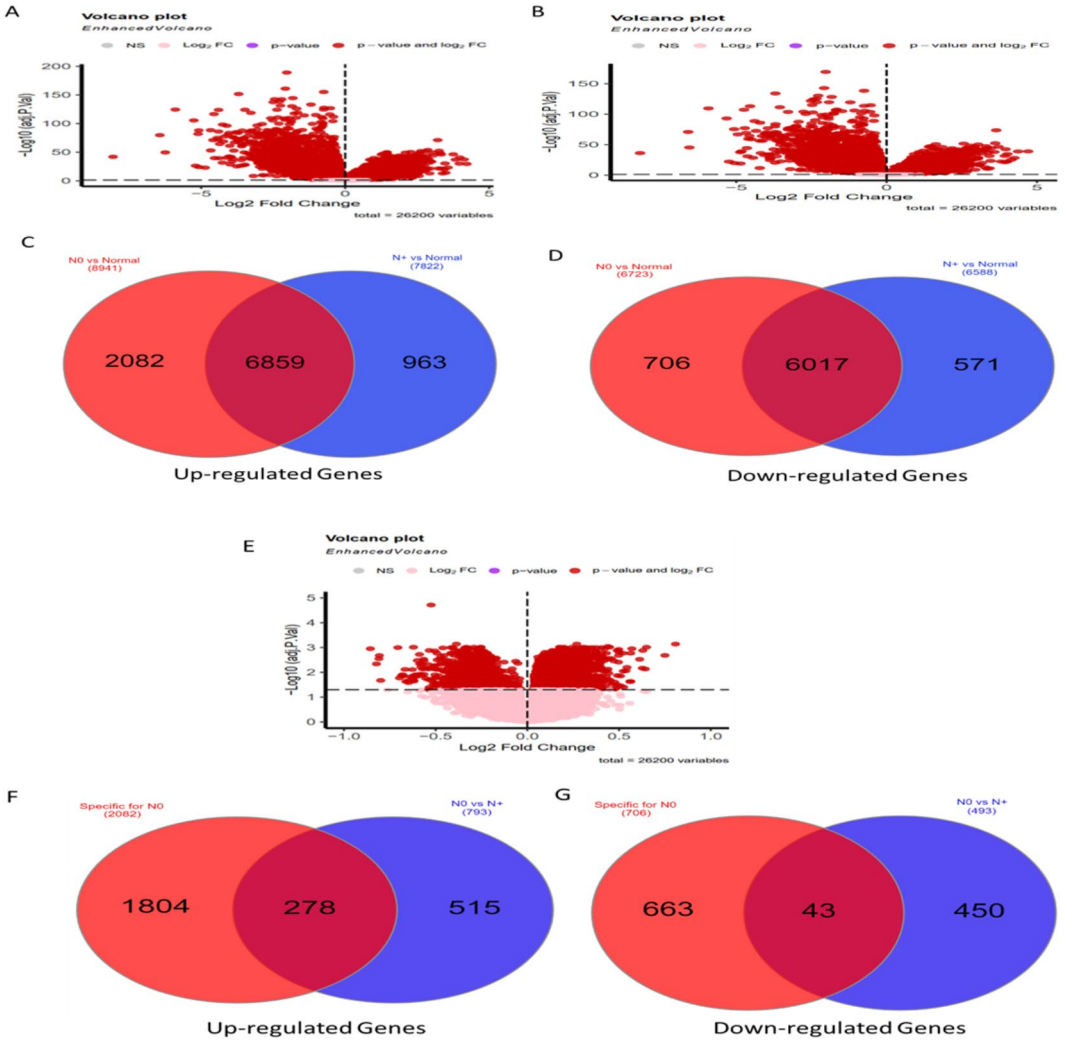

Fig. 2. Volcano plots of differentially expressed RNAs of LUAD patients from the TCGA were generated using the edgeR and ggplot packages. (**A**) N0 status vs. normal samples and (**B**) N + status vs. normal samples. (**C**) Venn diagrams show the 2082 genes upregulated in N0 status vs. normal, and (**D**) 706 genes downregulated in N0 status vs. normal. (**E**) N0 vs. N + status, difference expression of all genes in N0 compared to N+. (**F**) Venn diagrams showing the 278 genes upregulated in N0 vs. N + status compared to normal. (**G**) Venn diagrams showing the 43 genes downregulated in N0 vs. N + status compared to normal.

Diagnostic and prognostic value of the selected genes in N0- LUAD

To verify the diagnostic efficacy of selected lncRNAs and mRNAs specific for the N0 status of LUAD including LINC00894, YEATS2-AS1, and SUGP2 genes we analyzed their experimental expressions of patients using receiver operating characteristic (ROC) curves. The area under the ROC curve (AUC) is an important indicator that reflects both sensitivity and specificity of biomarkers for diagnostic tests. AUC values range from 0 to 1 and AUC greater than 0.5 is considered better than random chance and demonstrates some ability to differentiate between individuals with and without a specific disease. An AUC close to 1 indicates excellent diagnostic performance, with a value of 1 signifying perfect accuracy[18,19]. Our results showed that all three genes could serve as excellent biomarkers for the N0 status of LUAD, given that the AUC is closer to 1 (Fig. 6A-C).

Furthermore, bioinformatic survival analysis indicated that patients with higher expression of LINC00894, and SUGP2 had a significantly lower 5-year survival rate compared with patients with low expression of LINC00894, and SUGP2 (log-rank $P < 0.05$; Fig. 6D-E).

Knockdown of LINC00894 and its effect on the migration, invasion, and apoptosis

Given that LINC00894 had an intricate subnetwork with the most interactions with mRNAs and miRNAs, as well as being validated by experimental analyses as N0-specific DEGs and an excellent biomarker in our ROC curve analyses, we further explored the impact of LINC00894 knockdown using shRNA on the calu-3 cell line.

In the 24-hour experimental group, the expression level of LINC00894 was significantly decreased compared to the control group ($p = 0.03$) as well as shNC ($p = 0.006$) and Empty-vector control groups ($p = 0.0008$) (Fig. 7A). These results are similar in the 48-hour experimental groups, as well (Fig. 7B).

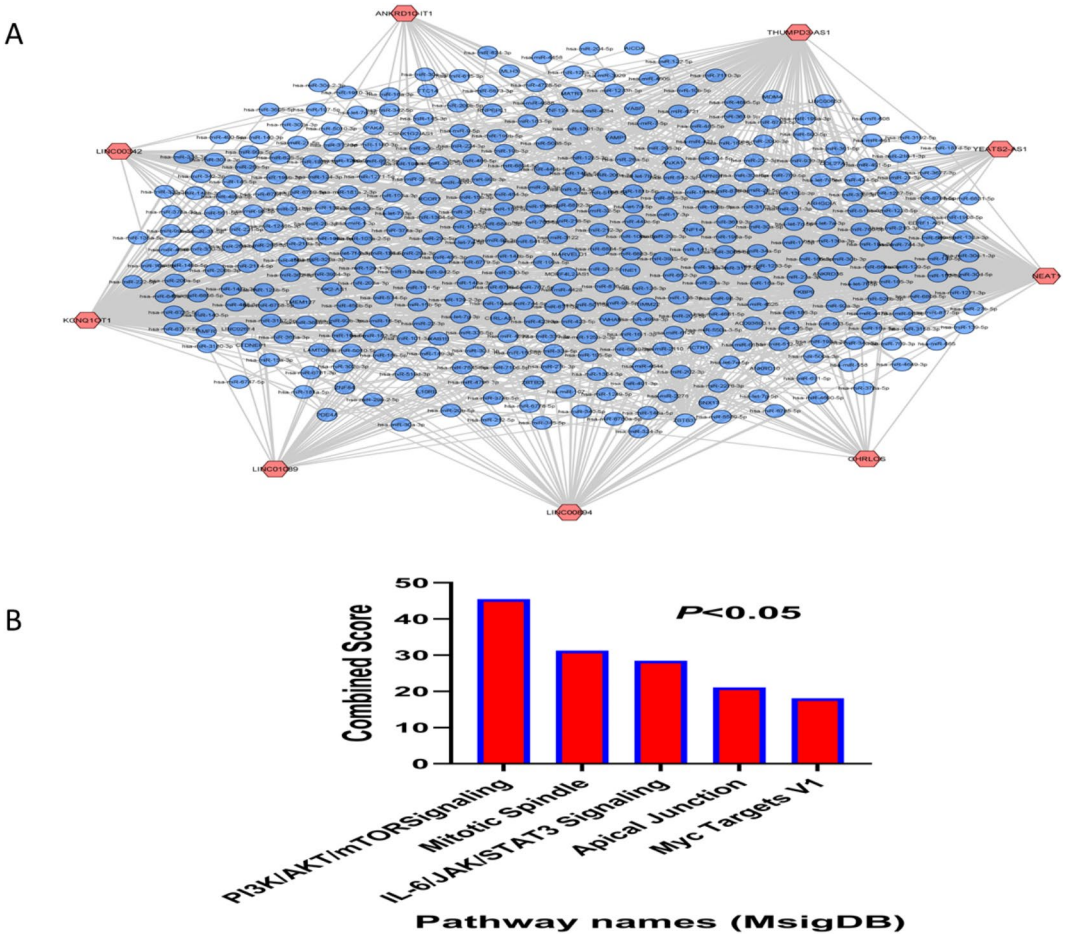

Fig. 3. The LUAD-specific lncRNA network and signaling pathways of all genes within the ceRNA network. (**A**) Blue represents lncRNA, mRNA, and miRNA seeds, while the orange hexagon represents the candidate lncRNA of interest. (**B**) Bioinformatic analysis of ceRNA network pathways.

Degree	LncRNA ID	Name	N0. vs. Normal logFold Chang (*p*-value)	N0. vs. N^+ log Fold Change (*p*-value)
25	ENSG00000229152	ANKRD10-IT1	0.73 (2.50E-09)	0.4 (0.001)
27	ENSG00000233885	YEATS2-AS1	0.38 (2.25E-07)	0.22 (0.003)
55	ENSG00000235703	LINC00894	0.65 (4.96E-07)	0.33 (0.008)
41	ENSG00000232931	LINC00342	1.25 (7.18E-08)	0.62 (0.006)
220	ENSG00000269821	KCNQ1OT1	0.96 (2.28E-07)	0.629 (0.001)
111	ENSG00000206573	THUMPD3-AS1	0.55 (5.94E-08)	0.27 (0.006)
381	ENSG00000245532	NEAT1	0.55(0.008126)	0.53 (0.008)
25	ENSG00000240288	GHRLOS	0.24(1.23E-05)	0.18 (0.001)
6	ENSG00000212694	LINC01089	0.46 (0.0003)	0.34 (0.007)

Table 2. Top-rank lncRNAs related to N0 status in LUAD.

Next, we assessed the effects of LINC00894 knockdown on the invasion and migration abilities of calu-3 cells. Our results showed that the invasion and migration ability of the shRNA-transfected cells significantly reduced compared to the control group ($p = 0.01$) only in the 48-hour group (Fig. 7C-D).

Furthermore, flow cytometry data using the Annexin V-FITS test indicated that in the 48-hour group, the early-stage apoptosis rate significantly increased from 0.50 to 23.07% in the shLINC00894(48 h) group ($p = 0.007$). however, there was no significant alteration in other conditions (Fig. 8A-H).

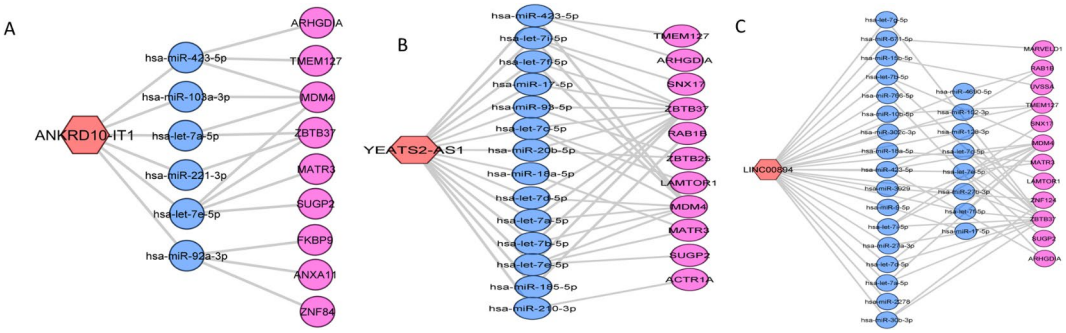

Fig. 4. The selected ceRNA sub-networks for deregulated genes specific to the N0 status of LUAD including (**A**) ANKRD10-IT1, (**B**) YEATS2-AS1, and (**C**) LINC00894 lncRNAs.

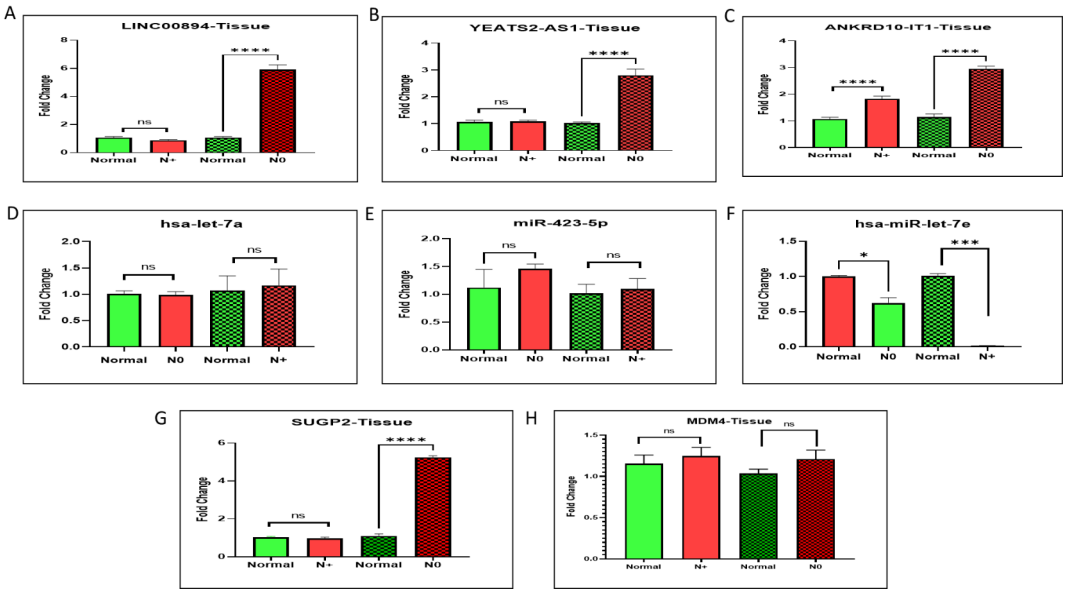

Fig. 5. Expression analyses of selected lncRNAs, miRNAs, and mRNAs in LUAD tissue samples including N0, N+, and adjacent normal tissues using qRT-PCR. GAPDH was used as an internal control for lncRNAs and mRNA, while U6 was used as an internal control for miRNAs. (**A**) LINC00894, (**B**) YEATS2-AS1 and (**G**) SUGP2 genes were significantly and specifically deregulated in N0 status but (**C**) ANKRD10-IT1, (**D**) hsa-let-7a-5p, (**E**) hsa-miR-423-5p, (**F**) hsa- let7e-5p, and (**H**) MDM4 genes were not significantly or specifically deregulated in N0 status.

Discussion

LUAD is a malignant disease with a poor prognosis and limited treatment options[20] especially in advanced stages[21]. Hence, it is crucial to investigate new biomarkers to improve the early diagnosis of LUAD. In this study, we focused on the N0 status of LUAD (early stage).

In our study, using the TCGA database, we found 278 and 43 genes up- and down-regulated, respectively (totally including 122 lncRNAs, and 91 mRNAs, and the others were not annotated), which were not only deregulated in N0 vs. normal but also in N0 vs. N + in LUAD patients. Most of them (122 genes), were classified as lncRNAs, and considering that lncRNAs compete with mRNA for miRNAs binding, we obtained a ceRNA network and found that there were nine lncRNAs with the highest degree.

Our analyses indicated that PI3K/AKT/mTOR pathway had the highest score among the genes in the ceRNA network. In line with our results, it was reported that in non-small cell lung cancer (NSCLC) including LUAD, deregulation of the PI3K/Akt/mTOR pathway by mutations or up and down-regulation of components of this cascade has been profoundly involved in tumorigenesis and the progression of disease[22] as increased p-mTOR was seen in up to 90% of patients with adenocarcinoma, and also downstream products of mTOR activation, including S6K and 4E-BP1 were observed in NSCLC with a more predominance in adenocarcinoma[23]. However, some studies have shown that components of the PI3K/Akt/mTOR pathway are involved in the early as well as advanced stage of LUAD[24,25]. The PI3K pathway regulates a complex cascade of events that are crucial for cell survival and growth. It all begins with the recruitment of Akt by activated PI3K, followed by Akt phosphorylation

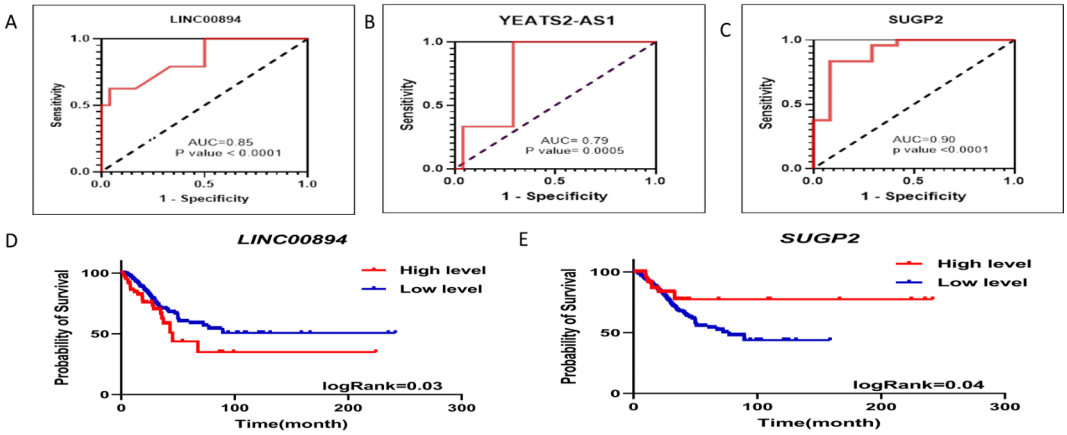

Fig. 6. The diagnostic and prognostic significance of three lncRNAs and SUGP2 in LUAD patients without lymph node metastasis. ROC curve analysis demonstrated that (**A**) LINC00894 (AUC = 0.85), (**B**) YEATS2-AS1 (AUC = 0.79), and (**C**) SUGP2 (AUC = 0.90) can be considered as diagnostic biomarkers. Survival analyses from the TCGA database indicated that increased expression levels of (**D**) LINC00894 and (**E**) SUGP2 are associated with a higher mortality rate in patients with adenocarcinoma (Log-rank test, $p < 0.05$).

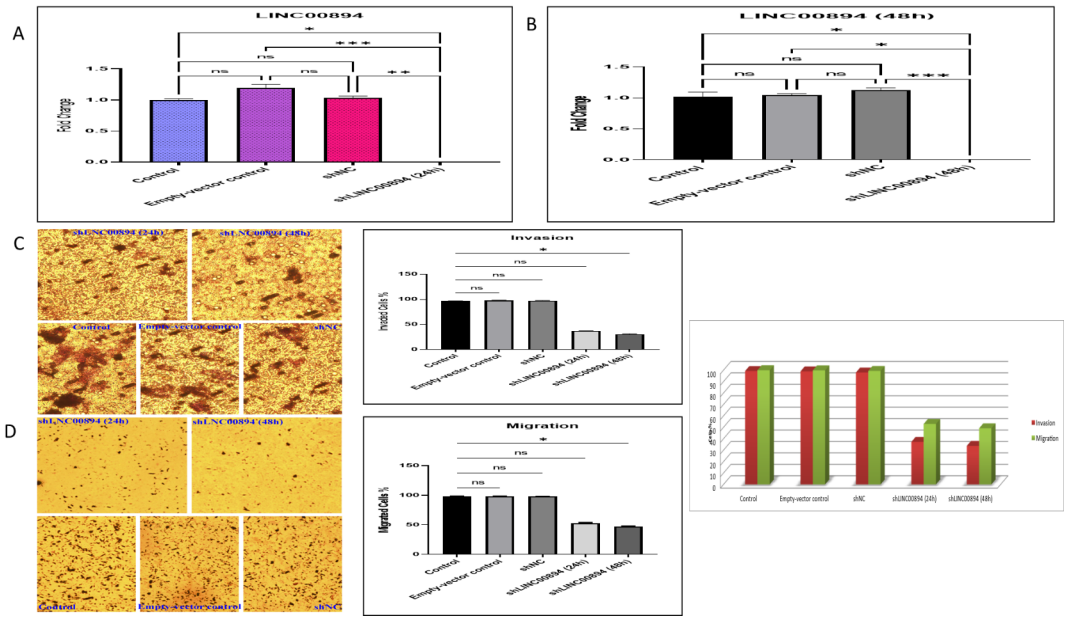

Fig. 7. Expression Assay of Knockdown LINC00894 in Calu-3 Cell Line After Transfection. LNC00894 knockdown in calu-3 cells. (**A**,**B**) Levels of LNC00894 in calu-3 cells infected with Control-Empty-vector control-shNC, and shLNC00894 were determined using qRT-PCR on day 5 post-infection. Note that the LNC00894 level was efficiently downregulated after shRNA infection. Knockdown of LINC00894 inhibits migration and invasion of the calu-3 cell line. Calu-3 cells were transfected with Control, Empty-vector, control- shNC, and shLNC00894 for 24 h and 48 h. The relative expression of LINC00894 was detected by RT-qPCR. (**C**) Invasion and (**D**) migration capacities of calu-3 cells were analyzed using Transwell assays.

by PDK, leading to Akt activation. Once activated, Akt exerts its effects by inhibiting pro-apoptotic proteins like BAD and BAX, thereby promoting cell survival. Moreover, Akt stimulates the activity of key players like GSK-3β and mTORC1,which are essential for cell growth, metabolism, and protein synthesis[22].

In the context of lung cells, Akt activation has been closely linked to the initiation and progression of cancer, especially in response to exposure to tobacco carcinogens. The activation of the PI3K/Akt pathway in lung epithelial cells can be triggered by various upstream signals, including Ras, growth factor tyrosine kinase receptors, and the tumor suppressor PTEN. Among these signals, the epidermal growth factor receptor (EGFR) stands out as a primary upstream activator of the PI3K/Akt pathway[26]. It is therefore reasonable that this pathway plays a key role in primary LUAD without lymph node metastasis.

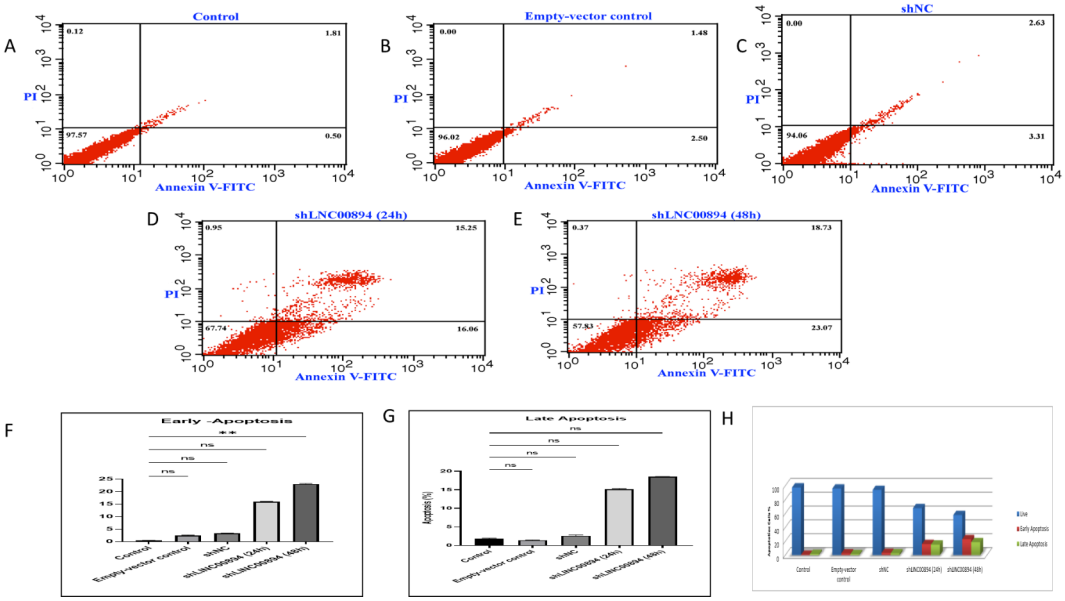

Fig. 8. The effect of LINC00894 knockdown on apoptosis induction in the Calu-3 cell line was determined by Annexin V staining followed by flow cytometric analysis. (**A**-**E**) Flow cytometry data showed that in the 48-hour group, the early-stage apoptosis rate was significantly increased in the shLINC00894 group. (**G**) Bar chart showing the percentage of apoptotic cells.

We also performed subnetwork analyses among the ceRNA network to gain more informative insights into the role of the identified genes. Among them, we selected three subnetworks with the most interactions, which include ANKRD10-IT1, YEATS2-AS1, and LINC00894-associated ceRNAs and assessed the expression of all three lncRNAs, as well as three miRNAs including let7e-5p, miR-423-5p, and let-7a-5p, and two mRNAs including SUGP2 and MDM4, shared in all three subnetworks in the LUAD tissues.

Our results from bioinformatic and experimental showed that upregulation of the SUGP2, YEATS2-AS1, and LINC00894 genes are specifically in the N0 status of LUAD, but in our experimental results, the ANKRD10-IT1 gene is not specifically deregulated and there is no MDM4 gene upregulation in the N0 LUAD condition, which is different from our bioinformatic results. We obtained our data for bioinformatic analyses from TCGA, which used RNA sequencing for gene expression profiling, a powerful tool for analyzing gene expression patterns on a large scale. However, despite its comprehensive nature, RNA sequencing may have limitations in terms of sensitivity and specificity when quantifying gene expression levels, leading to inconsistencies between experimental and bioinformatic results. In addition, insufficient sample sizes in experimental studies potentially resulting in conflicting results compared to bioinformatic analyses that utilize larger sample sizes.

Also, ROC curve analysis showed that SUGP2, YEATS2-AS1, and LINC00894 could be excellent biomarkers. Finally, our results showed that knockdown of the LINC00894 gene leads to decreased invasion and migration but increased apoptosis.

There are limited studies on the role of the SUGP2 gene in diseases including cancers. SUGP2 (SFRS14 or SRFS14) encodes a splicing factor involved in mRNA splicing. Deregulation of splicing factors involved in hematologic malignancies and solid tumors and have oncogenic or tumor-suppressive functions of various splicing factors[27]. It has been reported that SUGP2 is one of 51 identified genes for tumor recurrence in stage I of NSCLC including LUAD. Interestingly, in stage I tumors usually do not spread to lymph nodes, which is consistent with our results in N0 status of LUAD[28]. In addition, it was shown that mutations in SUGP2 can play a role in bladder cancer[29] and it was reported as a fusion gene with MLL-ELL-SFRS14 (SUGP2) in leukemia[30]. It is reasonable that SUGP2 deregulation could lead to abnormal mRNA processing that may contribute to tumorigenesis.

In our study, we found that lncRNAs play a key role in the N0 status of LUAD, particularly YEATS2-AS1, and LINC00894 were dysregulated both in bioinformatic and experimental analyses. In line with our results, using bioinformatic analyses, Zhong et al. recently reported that lncRNAs play a key role in LUAD[31].

Emerging evidence indicates that lncRNAs control various physiological and pathological processes[32,33], particularly in cancers, and can act as either oncogenes or tumor suppressor genes in LUAD[34,35]. Tissue-specific expression, high relative stability, unique expression profiles, detecting using simple techniques such as PCR, presence in biological fluids (including blood, saliva, and urine), enclosed in extracellular vehicles released by tumor cells (protection mechanism for degradation in fluids), and effect on the biological process by diverse mechanisms at epigenetic, transcriptional and posttranscriptional levels, leads to lncRNAs having the potential to be excellent candidate biomarkers. Also, they can be an effective RNA-based target for drugs including antisense oligonucleotides[36–38].

There are no studies on the role of YEATS2-AS1 lncRNA in LUAD, but bioinformatic analyses have shown that YEATS2-AS1 is upregulated in various cancers[39]. In line with our study, YEATS2-AS1 expression level

was significantly higher in soft tissue sarcomas (STS) and correlated with overall survival of patients and could be used as prognostic biomarkers in STS[40], endometrial[41], and prostate cancers[42]. Regarding to YEATS2-AS1 ceRNA subnetwork obtained from our bioinformatic results, which includes 14 miRNAs and mRNAs, one or more of these axes may be involved in N0 LUAD.

Furthermore, we found that LINC00894 is overexpressed in the N0 status of LUAD. In line with our results, in hepatocellular carcinoma, LINC00894 has been reported to be overexpressed in oxaliplatin-resistant cells[43] and kidney renal clear cell carcinoma[44], but in thyroid[45] and breast[46] cancers, its expression is downregulated and associated with poor prognosis. Interestingly, in papillary thyroid cancer (PTC), it was downregulated in the lymph node metastasis status[47], and given our finding that LINC00894 is overexpressed especially in the N0 status of LUAD, it may be involved in the N0 condition of LUAD and PTC. In addition, it was reported that LINC00894 is involved in tumorigenesis in the thyroid through acting as a let-7e-5p sponge[45]. It is reasonable that in LUAD as well, LINC00894 sponge let-7e-5p is one of the LINC00894-related ceRNA miRNAs from our bioinformatic results and leads to upregulation of downstream targets such as SUGP2 that we found that to be upregulated in this study by bioinformatic and experimental analyses. In addition, we found that the knockdown of LINC00894 results in decreased migration and invasion capabilities, while concurrently promoting apoptosis. In this context, our work hypothesizes that LINC00894 might act as an oncogene in lung adenocarcinoma.

Although the plasma membrane is intact in early apoptosis, phosphatidylserine is deposited on the cell surface. However, late apoptosis is also referred to as secondary necrosis, characterized by permeabilized plasma membrane through direct exposure of cells to extreme temperature, mechanical, and chemical insults, or prolongation of early apoptosis. Phagocytosis of early apoptotic cells leads to an anti-inflammatory response and can also promote cell growth and wound healing, but not in late apoptosis[48].

Little is known about biological the function and molecular mechanism or signaling pathways of SUGP2, LINC00894, and YEATS2-AS1 genes specific to the N0 status of LUAD identified in this study. In particular, LINC00894 and YEATS2-AS1 as lncRNAs could act through different mechanisms, including ceRNA that we obtained associated subnetwork in this study. In the future, the lncRNA-miRNA-mRNA axis of all axes obtained for both lncRNA can be experimentally assessed in LUAD and other cancers, as well as a precise mechanism of SUGP2, in early stage or N0 status of cancers. Also, in vivo experiments can be used for more validation of our results. The limitation of our study was that our bioinformatic data were only extracted from the TCGA database.

In conclusion, our bioinformatic and experimental analyses revealed that upregulation of LINC00894, YEATS2-AS1, and SUGP2 was involved especially in the N0 status of LUAD and they act as biomarker or prognostic markers for the N0 status of LUAD.

Data availability

The study extensively relied on publicly accessible datasets from TCGA http://gdac.broadinstitute.org/. All data generated or analyzed in the course of this study have been provided in the published article. Further datasets or additional data utilized in this study can be obtained from the corresponding author upon reasonable request.

Received: 7 March 2024; Accepted: 25 December 2024
Published online: 27 March 2025

References

1. Siegel, R. L., Miller, K. D., Fuchs, H. E. & Jemal, A. Cancer statistics. *CA Cancer J. Clin.* **72**(1), 7–33 (2022).
2. Chen, M., Liu, X., Du, J., Wang, X.-J. & Xia, L. Differentiated regulation of immune-response related genes between LUAD and LUSC subtypes of lung cancers. *Oncotarget* **8**(1), 133 (2017).
3. Bray, F. et al. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *Cancer J. Clin.* **68**(6), 394–424 (2018).
4. Cao, M. & Chen, W. Epidemiology of lung cancer in China. *Thorac. Cancer* **10**(1), 3–7 (2019).
5. Siegel, R. L., Miller, K. D. & Jemal, A. Cancer statistics, 2018. *Cancer J. Clin.* **68**(1), 7–30 (2018).
6. Marmor, H. N., Zorn, J. T., Deppen, S. A., Massion, P. P. & Grogan, E. L. Biomarkers in lung cancer screening: A narrative review. *Curr. Challenges Thorac. Surg.* **5**, 1–13 (2023).
7. Zhang, Z. et al. Sublobar resection is associated with better perioperative outcomes in elderly patients with clinical stage I non-small cell lung cancer: A multicenter retrospective cohort study. *J. Thorac. Disease* **11**(5), 1838 (2019).
8. Yerokun, B. A. et al. A national analysis of wedge resection versus stereotactic body radiation therapy for stage IA non–small cell lung cancer. *J. Thorac. Cardiovasc. Surg.* **154**(2), 675–686 (2017). e4.
9. Chen, Y., Jin, L., Jiang, Z., Liu, S. & Feng, W. Identifying and validating potential biomarkers of early stage lung adenocarcinoma diagnosis and prognosis. *Front. Oncol.* **11**, 644426 (2021).
10. Navani, N. et al. Lung cancer diagnosis and staging with endobronchial ultrasound-guided transbronchial needle aspiration compared with conventional approaches: An open-label, pragmatic, randomised controlled trial. *Lancet Respiratory Med.* **3**(4), 282–289 (2015).
11. Chen, Y., Zitello, E., Guo, R. & Deng, Y. The function of LncRNAs and their role in the prediction, diagnosis, and prognosis of lung cancer. *Clin. Transl. Med.* **11**(4), e367 (2021).
12. Wang, L. et al. 18F-FDG PET-based radiomics model for predicting occult lymph node metastasis in clinical N0 solid lung adenocarcinoma. *Quant. Imaging Med. Surg.* **11**(1), 215 (2021).
13. Khashei Varnamkhasti, K., Moghanibashi, M. & Naeimi, S. Genes whose expressions in the primary lung squamous cell carcinoma are able to accurately predict the progression of metastasis through lymphatic system, inferred from a bioinformatics analyses. *Sci. Rep.* **13**(1), 6733 (2023).
14. Wu, Y. et al. Driver and novel genes correlated with metastasis of non-small cell lung cancer: A comprehensive analysis. *Pathol. Res. Pract.* **224**, 153551 (2021).
15. Bridges, M. C., Daulagala, A. C. & Kourtidis, A. LNCcation: lncRNA localization and function. *J. Cell Biol.* **220**(2), e202009045 (2021).
16. Hu, Y., Gu, X., Duan, Y., Shen, Y. & Xie, X. Bioinformatics analysis of prognosis-related long non-coding RNAs in invasive breast carcinoma. *Oncol. Lett.* **20**(1), 113–122 (2020).

17. Akond, Z., Alam, M. & Mollah, M. N. H. Biomarker identification from RNA-seq data using a robust statistical approach. *Bioinformation* **14**(4), 153 (2018).
18. Obuchowski, N. A. Receiver operating characteristic curves and their use in radiology. *Radiology* **229**(1), 3–8 (2003).
19. Zhou, X.-H., Obuchowski, N. A. & McClish, D. K. *Statistical Methods in Diagnostic Medicine* (Wiley, 2014).
20. Zhou, Y. et al. Implications of different cell death patterns for prognosis and immunity in lung adenocarcinoma. *NPJ Precision Oncol.* **7**(1), 121 (2023).
21. Kerr, K. M. Pulmonary adenocarcinomas: Classification and reporting. *Histopathology* **54**(1), 12–27 (2009).
22. Sanaei, M-J., Razi, S., Pourbagheri-Sigaroodi, A. & Bashash, D. The PI3K/Akt/mTOR pathway in lung cancer; oncogenic alterations, therapeutic opportunities, challenges, and a glance at the application of nanoparticles. *Transl. Oncol.* **18**, 101364 (2022).
23. Tan, A. C. Targeting the PI3K/Akt/mTOR pathway in non-small cell lung cancer (NSCLC). *Thorac. Cancer* **11**(3), 511–518 (2020).
24. Krencz, I. et al. Expression of mTORC1/2-related proteins in primary and brain metastatic lung adenocarcinoma. *Hum. Pathol.* **62**, 66–73 (2017).
25. Xu, L. et al. Single-cell RNA sequencing reveals the mechanism of PI3K/AKT/mTOR signaling pathway activation in lung adenocarcinoma by KRAS mutation. *J. Gene Med.* **26**(1), e3658 (2024).
26. Ghareghomi, S., Atabaki, V., Abdollahzadeh, N., Ahmadian, S. & Ghoran, S. H. Bioactive PI3-kinase/Akt/mTOR inhibitors in targeted lung cancer therapy. *Adv. Pharm. Bull.* **13**(1), 24 (2021).
27. Yan, Y., Ren, Y., Bao, Y. & Wang, Y. RNA splicing alterations in lung cancer pathogenesis and therapy. *Cancer Pathogenesis Therapy* **1**(04), 272–283 (2023).
28. Lu, Y., Wang, L., Liu, P., Yang, P. & You, M. Gene-expression signature predicts postoperative recurrence in stage I non-small cell lung cancer patients. *PLoS ONE* **7**(1), e30880 (2012).
29. Zhao, J. et al. Whole-exome sequencing of muscle-invasive bladder cancer identifies recurrent copy number variation in IPO11 and prognostic significance of importin-11 overexpression on poor survival. *Oncotarget* **7**(46), 75648 (2016).
30. Grossmann, V. et al. Targeted next-generation sequencing detects point mutations, insertions, deletions and balanced chromosomal rearrangements as well as identifies novel leukemia-specific fusion genes in a single procedure. *Leukemia* **25**(4), 671–680 (2011).
31. Zhong, J. et al. Identification and functional characterization of PI3K/Akt/mTOR pathway-related lncRNAs in lung adenocarcinoma: A retrospective study. *Cell. J. (Yakhteh)* **26**(1), 13 (2024).
32. Beermann, J., Piccoli, M.-T., Viereck, J. & Thum, T. Non-coding RNAs in development and disease: Background, mechanisms, and therapeutic approaches. *Physiol. Rev.* **96**(4), 1297–1325 (2016).
33. Nadhan, R. & Dhanasekaran, D. N. Decoding the oncogenic signals from the long non-coding RNAs. *Onco* **1**(2), 176–206 (2021).
34. Ghahramani Almanghadim, H. et al. New insights into the importance of long non-coding RNAs in lung cancer: Future clinical approaches. *DNA Cell Biol.* **40**(12), 1476–1494 (2021).
35. Gencel-Augusto, J., Wu, W. & Bivona, T. G. Long non-coding RNAs as emerging targets in lung cancer. *Cancers* **15**(12), 3135 (2023).
36. Zhang, Y. & Tang, L. The application of lncRNAs in cancer treatment and diagnosis. *Recent Pat. Anti-cancer Drug Discov.* **13**(3), 292–301 (2018).
37. Le, P., Romano, G., Nana-Sinkam, P. & Acunzo, M. Non-coding RNAs in cancer diagnosis and therapy: Focus on lung cancer. *Cancers* **13**(6), 1372 (2021).
38. Ratti, M. et al. MicroRNAs (miRNAs) and long non-coding RNAs (lncRNAs) as new tools for cancer therapy: First steps from bench to bedside. *Target. Oncol.* **15**, 261–278 (2020).
39. Dai, C., Zeng, X., Zhang, X., Liu, Z. & Cheng, S. Machine learning-based integration develops a mitophagy-related lncRNA signature for predicting the progression of prostate cancer: A bioinformatic analysis. *Discover Oncol.* **15**(1), 316 (2024).
40. Zhang, X. et al. N6-Methyladenosine-related lncRNAs are anticipated biomarkers for sarcoma patients. *J. Oncol.* **2022**(1), 1093805 (2022).
41. Zhou, X., Zhang, H., Duan, Y., Zhu, J. & Dai, H. m6A-related long noncoding RNAs predict prognosis and indicate therapeutic response in endometrial carcinoma. *J. Clin. Lab. Anal.* **37**(1), e24813 (2023).
42. Zhang, P., Tan, X., Zhang, D., Gong, Q. & Zhang, X. Development and validation of a set of novel and robust 4-lncRNA-based nomogram predicting prostate cancer survival by bioinformatics analysis. *PLoS ONE* **16**(5), e0249951 (2021).
43. Huang, J. et al. LINC00894 targets annexin A2 to regulate oxaliplatin resistance in hepatocellular carcinoma: ANXA2 protein function. *Int. J. Biol. Macromol.* **281**, 136538 (2024).
44. Deng, L., Wang, P., Qu, Z. & Liu, N. The construction and analysis of ceRNA network and immune infiltration in kidney renal clear cell carcinoma. *Front. Genet.* **12**, 667610 (2021).
45. Chen, B., Liu, D., Chen, R., Guo, L. & Ran, J. Elevated LINC00894 relieves the oncogenic properties of thyroid cancer cell by sponging let-7e-5p to promote TIA-1 expression. *Discover Oncol.* **13**(1), 56 (2022).
46. Zhang, X., Wang, M., Sun, H., Zhu, T. & Wang, X. Downregulation of LINC00894-002 contributes to tamoxifen resistance by enhancing the TGF-β signaling pathway. *Biochemistry (Moscow)* **83**, 603–611 (2018).
47. Zhou, X. et al. The m6A methyltransferase METTL3 drives thyroid cancer progression and lymph node metastasis by targeting LINC00894. *Cancer Cell Int.* **24**(1), 47 (2024).
48. Poon, I. K. H., Hulett, M. D. & Parish, C. R. Molecular mechanisms of late apoptotic/necrotic cell clearance. *Cell Death Differ.* **17**(3), 381–397 (2010).

Author contributions

"M.M. designed the study and critically reviewed the manuscript. M.A. performed all analyses. P.M. and S.N. critically reviewed the manuscript and interpreted the data. The final manuscript has been approved by all authors."

Funding

The authors declare that they did not receive any funding, grants, or other support during the preparation of this manuscript.

Declarations

Competing interests

The authors declare no competing interests.

Ethics approval

The research protocol was approved by the Ethics Committee of the Islamic Azad University, Kazarun branch (IR.IAU.KAU.REC.1400.142). All methods were performed in accordance with the relevant guidelines and regulations.

Consent to participate

By signing the consent form, the patients agreed to participate in the research and authorized the use of their samples for research purposes.

Additional information

Correspondence and requests for materials should be addressed to M.M. or S.N.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.