Translational Oncology 27 (2023) 101571

Contents lists available at ScienceDirect

Translational Oncology

journal homepage: www.elsevier.com/locate/tranon

Single-cell RNA-seq analysis to identify potential biomarkers for diagnosis, and prognosis of non-small cell lung cancer by using comprehensive bioinformatics approaches

Adiba Sultana [a,b,c,#], Md Shahin Alam [a,#], Xingyun Liu [c], Rohit Sharma [d,*], Rajeev K. Singla [c,e,*], Rohit Gundamaraju [f], Bairong Shen [c,*]

[a] School of Biology and Basic Medical Sciences, Soochow University Medical College, 199 Ren'ai Road, Suzhou 215123, China
[b] Center for Systems Biology, Soochow University, Suzhou 215006, China
[c] Institutes for Systems Genetics, Frontiers Science Center for Disease-Related Molecular Network, West China Hospital, Sichuan University, Xinchuan Road 2222, Chengdu, Sichuan, China
[d] Department of Rasa Shastra and Bhaishajya Kalpana, Faculty of Ayurveda, Institute of Medical Sciences, Banaras Hindu University, Varanasi, Uttar Pradesh 221005, India
[e] School of Pharmaceutical Sciences, Lovely Professional University, Phagwara, Punjab 144411, India
[f] ER Stress and Mucosal Immunology Lab, School of Health Sciences, College of Health and Medicine, University of Tasmania, Launceston, Tasmania, TAS 7248, Australia

ARTICLE INFO

Keywords:
Non-small cell lung cancer
Single-cell RNA-sequencing
Networking analysis
Gene biomarkers
Diagnosis

ABSTRACT

Non-small cell lung cancer (NSCLC) is the most common type of lung cancer and the leading cause of cancer-related deaths worldwide. Identification of gene biomarkers and their regulatory factors and signaling pathways is very essential to reveal the molecular mechanisms of NSCLC initiation and progression. Thus, the goal of this study is to identify gene biomarkers for NSCLC diagnosis and prognosis by using scRNA-seq data through bioinformatics techniques. scRNA-seq data were obtained from the GEO database to identify DEGs. A total of 158 DEGs (including 48 upregulated and 110 downregulated) were detected after gene integration. Gene Ontology enrichment and KEGG pathway analysis of DEGs were performed by FunRich software. A PPI network of DEGs was then constructed using the STRING database and visualized by Cytoscape software. We identified 12 key genes (KGs) including *MS4A1, CCL5*, and *GZMB*, by using two topological methods based on the PPI networking results. The diagnostic, expression, and prognostic potentials of the identified 12 key genes were assessed using the receiver operating characteristics (ROC) curve and a web-based tool, SurvExpress. From the regulatory network analysis, we extracted the 7 key transcription factors (TFs) (FOXC1, YY1, CEBPB, TFAP2A, SREBF2, RELA, and GATA2), and 8 key miRNAs (hsa-miR-124-3p, hsa-miR-34a-5p, hsa-miR-21-5p, hsa-miR-155-5p, hsa-miR-449a, hsa-miR-24-3p, hsa-let-7b-5p, and hsa-miR-7-5p) associated with the KGs were evaluated. Functional enrichment and pathway analysis, survival analysis, ROC analysis, and regulatory network analysis highlighted crucial roles of the key genes. Our findings might play a significant role as candidate biomarkers in NSCLC diagnosis and prognosis.

Abbreviations

NSCLC	Non-Small Cell Lung Cancer
scRNA-seq	Single-cell RNA sequencing
GEO	Gene Expression Omnibus
DEGs	Differentially expressed genes
GO	Gene Ontology
KEGG	Kyoto Encyclopedia of Genes and Genomes
PPI	Protein-protein interaction
KGs	Key Genes
ROC	Receiver Operating Characteristics
TF	Transcription factors

* Corresponding authors.
E-mail addresses: rohitsharma@bhu.ac.in (R. Sharma), rajeevsingla26@gmail.com, rajeevkumar@scu.edu.cn (R.K. Singla), bairong.shen@scu.edu.cn (B. Shen).
# These authors contributed equally to this study.

https://doi.org/10.1016/j.tranon.2022.101571
Received 30 March 2022; Received in revised form 12 October 2022; Accepted 24 October 2022

LUSC	Lung squamous cell carcinoma
LUAD	Lung adenocarcinoma
GAPDH	Glyceraldehyde-3 phosphate dehydrogenase
t-SNE	t-distributed stochastic neighbor embedding
STRING	Search Tool for the Retrieval of Interacting Genes
AUC	Area under curve
logFC	Log of Fold Change
CI	Confidence Interval
SVM	Support Vector Machine
SCLC	Small Cell Lung Cancer

Introduction

Cancer, a heterogeneous disease, poses a serious challenge for precise treatment at the individual level. Both bulk and single-cell RNA sequencing (scRNA-seq) technologies are used for studying transcriptional profiles at the gene expression level. Several articles have identified biomarkers for NSCLC diagnosis using bulk RNAseq technology [1–5]. scRNA-sequencing categorizes the cell types across multiple tissues, whereas bulk RNA sequencing involves the use of a tissue or cell population [6,7]. scRNA-seq is widely used to determine tumor heterogeneity, cellular identities, novel biomarkers, and molecular and functional strategies [8]. Several scRNA-seq based studies have been performed earlier to explore tumor heterogeneity and to identify novel biomarkers for different cancers [9–12].

Non-small cell lung cancer (NSCLC) is a highly heterogeneous lung cancer, accounting for approximately 85% of all the types of lung cancers, and it is strongly correlated with smoking habits [13,14]. NSCLC is mainly classified into two groups: lung squamous cell carcinoma (LUSC) and lung adenocarcinoma (LUAD). Most of the current treatment strategies for NSCLC are chemotherapies based on the histology and targeted agents for patients [15,16]. Treatment outcomes of NSCLC are quite insufficient as the post-therapy relapse rate and drug resistance remains high, and the 5-year relative survival rate of the patients is 26% [17,18]. Further exploration into the underlying mechanisms of NSCLC is thus quite urgent, which will impact the discovery of novel diagnostics and provide effective & key targets for NSCLC.

Expanding research recommends that affluent genes, miRNAs, TFs, and/or biological enriched pathways are associated with the development and progression of cancers. The altered expression of miRNAs has a strong correlation with distinct disease and carcinoma [19]. Gene-miRNAs interactions have been generally exhibited to control the convoluted molecular systems' basic oncogenesis, advancement, intrusion, and tumor metastasis [20]. Several studies have recently shown a link of miRNA in cancer development [21]. Some studies have highlighted the importance of miRNA expression in lung tissues [22]. The deregulation of miRNA might play a role in fatal NSCLC progression [23]. Thus, several studies have identified some important TFs and miRNAs as transcriptional and post-transcriptional factors for cancer through the regulatory interaction networks [24,25].

In the last decade, several biomarkers have been reported as prognostic and predictive markers for NSCLC [2,26-30]. Song et al. identified *STAT3, EGFR, PTEN, KRAS, TP53, RHOA, CTNNB1,* and *VEGFA,* as efficient targeted genes associated with six miRNAs (hsa-miR-21-5p, hsa-miR-31-5p, hsa-miR-708-5p, hsa-miR-30a-5p, hsa-miR-451a, and has-miR-126-3p) by expression analysis and miRNA-hub gene network for NSCLC [2]. Chen et al. discovered four genes viz. CDK1, PLK1, RAD51, and RFC4 as novel biomarkers using microarray gene expression profiles that might be potential therapeutic targets in NSCLC [26]. Valk et al. identified *SPAG5, POLH, KIF23, RAD54L, SGCG, NLRC4, MMRN1*, and *SFTPD* as the novel genes that are involved in multiple pathways leading to NSCLC [31]. Puzone et al. showed that the overexpression of glyceraldehyde-3 phosphate dehydrogenase (*GAPDH*) correlates with poor prognosis in NSCLC patients [32]. These findings are important to understand NSCLC pathogenesis.

Although much research has already been conducted to reveal the molecular mechanism of NSCLC progression, the heterogeneity and complexity of NSCLC still poses a great challenge and need for novel and effective biomarkers. Recently, scRNA-seq technology has been used to detect tumor heterogeneity and explore the gene expression pattern in tissues that can help the researcher to detect the novel biomarkers. Here, we utilized computational models for analyzing the scRNA-seq data to reveal the tumor heterogeneity of NSCLC tissues. We identified differentially expressed genes (DEGs), their associated pathways, and PPI network to screen the key genes (KGs), key miRNAs, and key TFs for personalized diagnosis and prognosis of NSCLC and performed further analysis to validate the result. Thus, the identification of important genes, miRNAs, and TFs as well as the signaling pathways related to cancer via bioinformatics analysis, will provide worthy enlightenment in cancer research.

Materials and methods

scRNA-seq data collection and processing

The publicly available scRNA-seq data (GSE127471, data collected from the peripheral blood of a patient with NSCLC by Newman and the team) were downloaded from the gene expression omnibus (GEO) database [33]. The data were sequenced on Illumina NextSeq 500 (Human). Data processing was performed using the Seurat package V3.1.1 in R V3.6.1 [34]. For quality control check, we extracted genes with a minimum number of features 200 having non zero counts and a minimum number of cells as 3. The filtered data were normalized using log-transformation and was used for further analysis. We used two datasets associated with NSCLC from the GEO database with accession numbers GSE19188 and GSE75037 to assess the diagnostic performance of the identified KGs.

Clustering and DEG identification

For dimensionality reduction, we performed principal component analysis (PCA) on the scaled data. The t-distributed stochastic neighbor embedding (t-SNE) was used to demonstrate two-dimensional data by first 10 principal components. The cell cluster was identified using K-means clustering based on the original Louvain algorithm. We used the Seurat's VlnPlot function to determine the expression of acquainted marker genes to assign clusters. Moreover, we constructed the trajectory analysis to reveal the tendency curve of the eight clusters using "Monocle" package [35]. The R package Seurat was used to analyze DEGs with scRNA-seq data.

Functional enrichment and pathway analysis of DEGs

We used a stand-alone software tool FunRich (version 3.1.4) for gene ontology (GO) functional enrichment and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis of DEGs [36]. We considered the threshold P-value < 0.01 to obtain significant functional and pathway terms. KEGG and GO were used to annotate the enrichment analysis of bio-term classification processes and genes clusters as well as to impart a fantasy module for inference [37,38].

PPI network analysis of DEGs

Protein-protein interactions (PPIs) networks facilitate the analysis of pathogenic mechanisms and disease progression by providing knowledge on the molecular mechanism underlying cellular activity. In this study, we used the Search Tool for the Retrieval of Interacting Genes (STRING v11.0) database for constructing PPI network of DEGs [39]. Cytoscape (version 3.7.1) was used to discern the PPI networks between the DEGs [40]. We identify 12 KGs using two topological methods Betweenness and Stress in Cytoscape plugin *cytoHubba* [41].

Survival analysis and expression level of KGs

We used SurvExpress to check the impact of the expression pattern and survival analysis of KGs. SurvExpress (http://bioinformatica.mty.itesm.mx/SurvExpress) is an online tool for analyzing cancer gene expression data for the validation and survival analysis of multi-gene biomarkers [42]. Here we used it to verify and estimate the impact the expression pattern and prognostic value of KGs using the Kaplan-Meier curve and log-rank test.

ROC analysis of the KGs

Receiver operating characteristic (ROC) curve analysis was carried out to assess the true positive rate (Sensitivity) and false-positive rate (1− Specificity) of the identified KGs using the "pROC" package in R [43]. Area under curve (AUC) was determined and used to screen the ROC values.

Regulatory interaction network analysis of KGs

We constructed a regulatory interaction network (KGs-miRNAs and KGs-TFs) of 12 KGs using an online tool NetworkAnalyst 3.0 [44]. It is a bioinformatics tool that visualizes and deciphers the information in the association of network settings. This tool included three gene-miRNAs interaction databases TarBase [45], miRTarbase [46] and miRecords [47] and three gene-TFs interaction databases such as ENCODE [48], JASPAR [49], and ChEA [50]. We used TarBase and JASPAR databases for KGs-miRNAs and KGs-TFs interaction networks, respectively.

Results

Tumor heterogeneity and identification of DEGs

The scRNA-seq technology provided a good transcriptional detailing of cancer cells and gene expression in NSCLC patients. The 1803 cells were adopted for advanced analysis. The 1803 cells were analyzed and then classified into 8 separate clusters based on the identical gene set. Every cluster was separated by identical gene set. To make the segregation clearer among clusters, DEG analysis in cell types based on their covariance patterns and mean expression levels were evolved. We found that the identified marker gene sets were significant to ascertain cell types individually with high potentiality. Heatmap of the top ten genes showing high heterogeneity among clusters based on logFC has been illustrated in Fig. 1(A). The results of trajectory analysis showed the cells of only cluster1 may have a significant difference from other cells in NSCLC Fig. 1(B). We identified DEGs using the Wilcoxon rank sum test, based on the threshold adj.P.Val < 0.01 and logFC > 2 for up-regulated genes and adj.P.Val < 0.01 and logFC < -2 for down-regulated genes. A total of 158 DEGs, of which 48 were up-regulated and 110 down-regulated genes were identified. All DEGs (up- and down-regulated) are listed in Table 1.

Functional enrichment and pathway analysis of DEGs

For further investigation, the GO functional and KEGG pathway enrichment analyses associated with the DEGs was performed in FunRich software. The top GO (BP, CC, and MF) and KEGG enrichment functions/terms of DEGs are shown in Fig. 2. In BP, 47.3% genes were enriched with protein metabolism, 8.90% genes were enriched in immune response, and 9.5% genes were enriched with cell growth and/or maintenance. In CC, 54.4% of genes were enriched in cytosol and cytoplasm; more than 40% of genes were enriched in exosomes, nucleolus, and ribosome; and approximately 30% of genes were enriched with

Table 1
Differentially expressed genes (DEGs) in NSCLC.

Up-regulated DEGs	Down-regulated DEGs
TREML1, TAGLN2, ***LTB****, **CCL5****, FTL, MNDA, LST1, SPARC, AIF1, ACRBP, AHSP,* ***MS4A1****, CD79B, VCAN, ALAS2,* ***CD74****, CST3, FCN1, PRF1, KLRB1,* ***GZMB****, FGFBP2,* ***TYROBP****, CMC1, CTSS, NRGN, RGS18, MYL9, HBD, CLU, TUBB1, SDPR, HIST1H2AC, S100A12, GNG11, CD79A, RP11-1143G9.4, CA1, LYZ, S100A9, S100A8, HBA2, GNLY, HBB, IGLL5, PF4,* ***HBA1****, PPBP*	*MALAT1, RPS29, RPL39, MT-ND3, RPS27, RPS28, RPL37, RPL34, RPS26,* ***ACTB****, RPS21, RPL23A, RPL36, MT-ND2, PTMA,* ***EEF1A1****, MT-CO1, RPL28, RPL41, RPL38, RPL26, TMSB4X, HLA-C, RPL10, TMSB10, RPLP2, MT-CO3, MT-ND4, MT-ND1, RPL3,* ***RPS18****, RPL27A, HA2, S100A6, RPS2, RPL27, RPS25, RPL36A, MT-ATP6, RPS8, B2M, RPS15A, HLA-B, RPL17, RPS24, RPL23,* ***RPLP1****, MT-CO2, RPS19, RPL13A, RPL35A, MT-CYB, RPSA, RPS23, RPS3, RPL35, RPS10, RPS16, RPS15, RPS6, VIM, RPL32, RPS13, RPL13, S100A4, RPS4X, RPL30, RPL31, NBEAL1, TXNIP, TPT1, HLA-A, FAU, RPS7, NEAT1,* ***GAPDH****, RPS3A, RPSAP58, RPS9, RPL19, RPL12, RPL15, RPL9, RPL11, RPS14, RPL6, ACTG1, H3F3A, RPL7, RPL7A, GNB2L1, RPS20, RPL5, PFN1, RPL29, RPL18, NKG7, RPL24, RPL22, TOMM7, RPL14, PABPC1, BTG1, RPL21,CFL1, RPS11, SRGN, RPL10A, RPL18A, HNRNPA1*

N.B. Bold with star (*) indicates key genes (KGs).

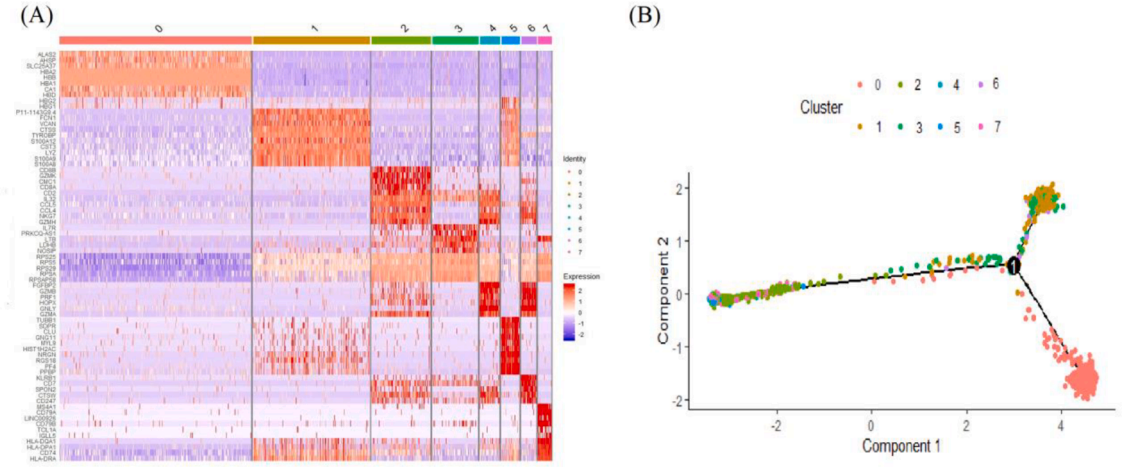

Fig. 1. (A) The heatmap of the top 10 marker genes of each cluster where each row represents genes and column represents clusters. (B) Trajectory plot differentiated by eight clusters.

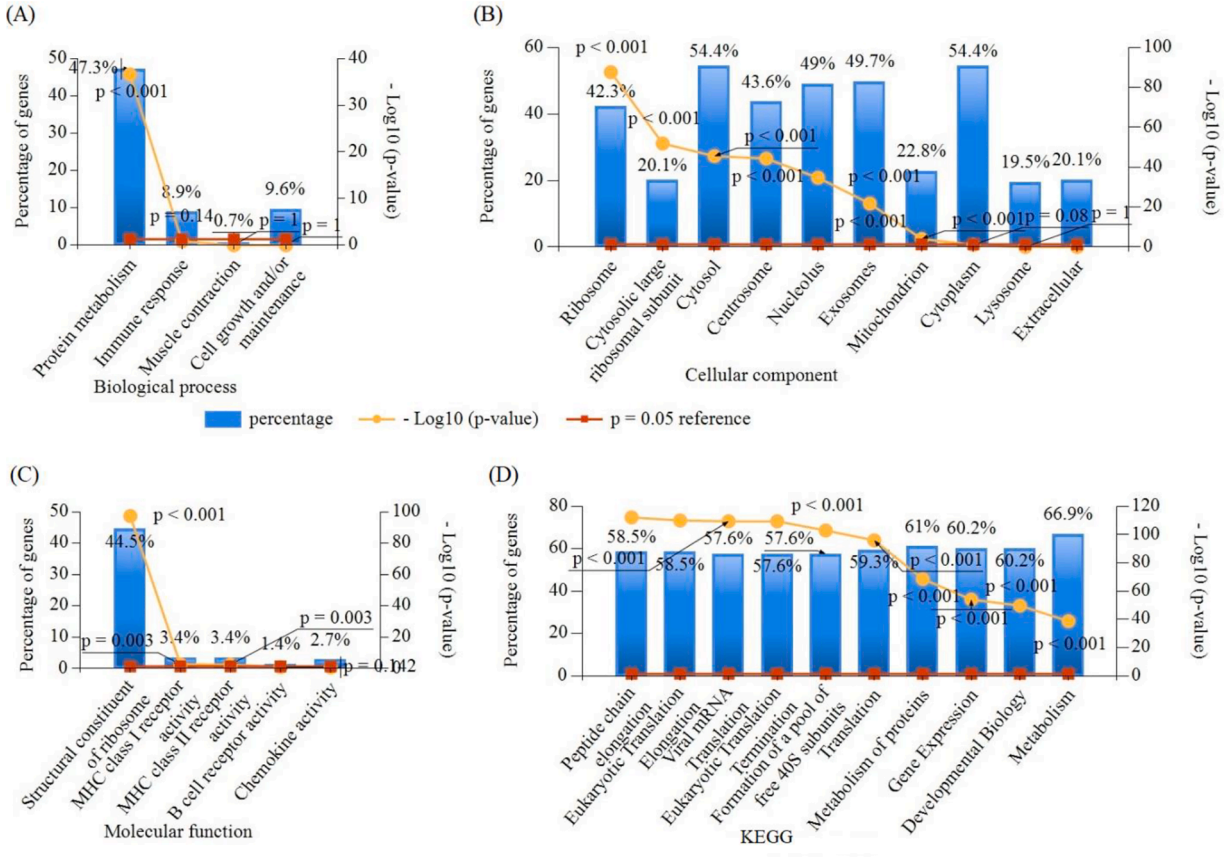

Fig. 2. Top GO and KEGG terms enriched by DEGs ((A) Biological processes, (B) Cellular components, (C) Molecular functions, and (D) Kyoto encyclopedia of genes and genome (KEGG)).

cytosolic large ribosomal subunit, extracellular, lysosome, etc., where the MF enrichment was mainly correlated with structural constituent of ribosome, MHC class I receptor activity, MHC class II receptor activity, B cell receptor activity, and chemokine activity terms in Table S1. The enriched KEGG pathways for the DEGs included peptide chain elongation, eukaryotic translation elongation, eukaryotic translation termination, viral mRNA translation, 3′ -UTR-mediated translational regulation, and metabolism of proteins pathways, which are associated with lung cancer development (Table S2).

PPI network of DEGs

Based on the online tool STRING, a total of 158 DEGs were used in the PPI network, involving 156 nodes and 3282 edges with an average node degree 42.2 and PPI enrichment P-value < 1.0e-16 (Fig. 3). The top 15 genes were selected using each of the two network scoring methods Betweeness and Stress in Cytoscape plugin, cytoHubba. We then extracted 12 common genes from the top 15 genes of the two methods and considered them as KGs, such as *MS4A1, CCL5, GZMB, HBA1, TYROBP, CD74, LTB, EEF1A1, RPLP1, RPS18, GAPDH,* and *ACTB,* as shown in Table 2. In addition, we assessed the significance of the 12 KGs through survival analysis, cancer prediction model, and regulatory interaction network analysis.

Survival analysis and expression level of KGs

The prognostic value of the identified KGs was evaluated by fitting Cox-proportional hazards regression model between high-risk and low risk group patients in Fig. 4(A), where red indicates high-risk group patients and green indicates low-risk group patients. We observed that the overall survival probability for the high-risk group compared to the low-risk group decreased over time based on the expression level of KGs (Hazard Ratio = 1.85 at 1.36-2.53 Confidence Interval (CI), and the log-Rank p-Value = 9.766e-05) This indicates that the proposed KGs have a strong prognostic power for NSCLC. The expression patterns of the KGs by the risk group are shown in the boxplot in Fig. 4(B). The overall results showed that the KGs displayed a significant prognostic performance for NSCLC that support our original results.

Performance measure of the identified KGs using ROC analysis

A supervised machine learning algorithm SVM classifier was considered to develop a Cancer Prediction Model for 12 KGs. We developed the cancer prediction model through the ROC curve for the training dataset with access number GSE19188 (red) and for the test dataset with access number GSE75037 (green) in Fig. 5. We observed that the AUC values range from 0.64 to 0.99 for the training dataset and 0.63 to 0.97 for the test dataset, which indicate the good prediction performance.

KGs-miRNAs and KGs-TFs interaction network analysis

To identify the key miRNAs associated with KGs, we constructed the interaction network between the KGs and miRNAs. We uploaded the official symbol in the gene list of human lung tissue and select the TarBase database in NetworkAnalyst to identify miRNAs that targeted the KGs (Fig. 6(A)). On the basis of degree score >= 4, we selected 8 key miRNAs (hsa-miR-124-3p, hsa-miR-34a-5p, hsa-miR-21-5p, hsa-miR-155-5p, hsa-miR-449a, hsa-miR-24-3p, hsa-let-7b-5p, and hsa-miR-7-5p) as post-transcriptional regulatory factors of KGs. From the literature review, we found that these miRNAs have a close relationship with drug resistance of lung cancer, basically NSCLC [51]. Similarly, the

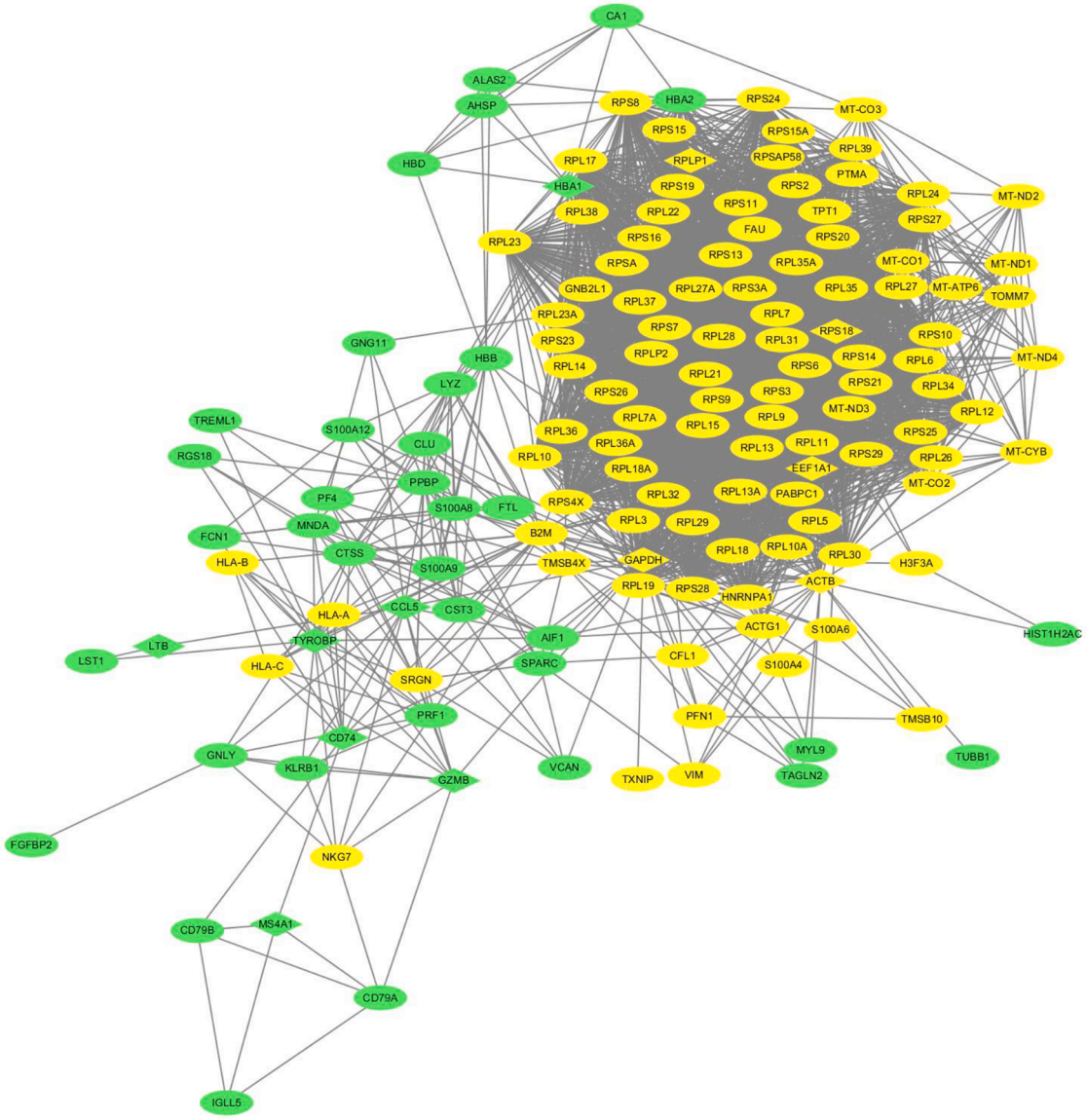

Fig. 3. Protein-protein interaction (PPI) network of DEGs. The green and yellow colors indicate up- and down-regulated genes respectively. The key genes (KGs) are highlighted in diamond shape.

Table 2
The summary of the key genes (KGs) identified from the Cytoscape in NSCLC.

S. N	Betweenness		Stress		Key Genes
	Name	Score	Name	Score	*MS4A1, CCL5, GZMB, HBA1, TYROBP, CD74, LTB, EEF1A1, RPLP1, RPS18, GAPDH, ACTB*
1	*GAPDH*	6144.813	*GAPDH*	146026	
2	*MS4A1*	2329.161	*MS4A1*	43470	
3	*CCL5*	1094.994	*CCL5*	23206	
4	*ACTB*	1056.283	*HBA1*	22566	
5	*GZMB*	830.074	*CD74*	20158	
6	*HBA1*	663.953	*GZMB*	20044	
7	*TYROBP*	662.6554	*ACTB*	17804	
8	*CD74*	634.4065	*LTB*	14866	
9	*HBA2*	551.3868	*TYROBP*	12728	
10	*EEF1A1*	474.6351	*EEF1A1*	11602	
11	*HBB*	462.9838	*RPS18*	11596	
12	*RPS3*	433.276	*RPLP1*	10574	
13	*RPLP1*	407.0451	*RPL13A*	10386	
14	*LTB*	399.4331	*PPBP*	10344	
15	*RPS18*	378.9975	*RPL19*	10194	

interaction network of KGs and TFs was constructed from the JASPAR database in NetworkAnalyst. Seven TFs (FOXC1, YY1, CEBPB, TFAP2A, SREBF2, RELA, and GATA2) with degree >= 4 were selected as transcriptional regulatory factors of KGs (Fig. 6(B)). Key miRNAs and TFs with their degree scores are listed in Table 3.

Discussion

NSCLC is a disease with a very high concern and is life-threatening for humans. Due to the heterogeneity of NSCLC, its treatment is very challenging. Hence, it will be helpful if NSCLC is managed by targeted treatment; however, patients with NSCLC have a lower prognosis. Hence, the identification of novel biomarkers based on the heterogeneity and the scRNA-seq data are one of the key tasks to improve the personalized and targeted medicine of NSCLC in the future.

In this study, we analyzed scRNA-seq data from the tumor tissue of peripheral blood of NSCLC patients to bioinformatically explore the cellular heterogeneity, DEGs, associated biological pathways, PPI network, key genes, miRNAs and TFs. The 1803 cells were classified into 8 explicit clusters while each cluster was mixed up with variant numbers

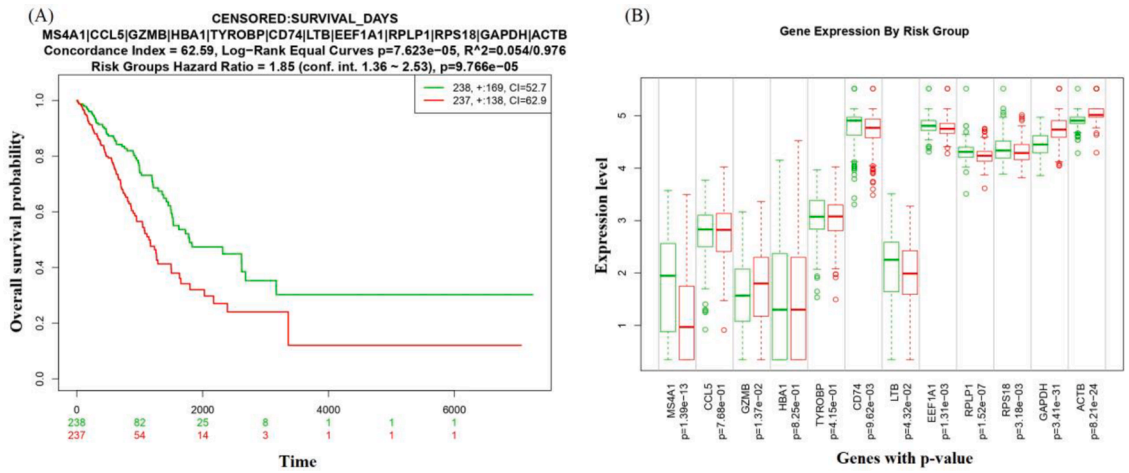

Fig. 4. (A) Kaplan-Meier plot displaying the prognostic effect of the KGs on NSCLC. (B) Boxplot displaying the expression pattern of KGs between risk groups. Red indicates high-risk group and green indicates low-risk group.

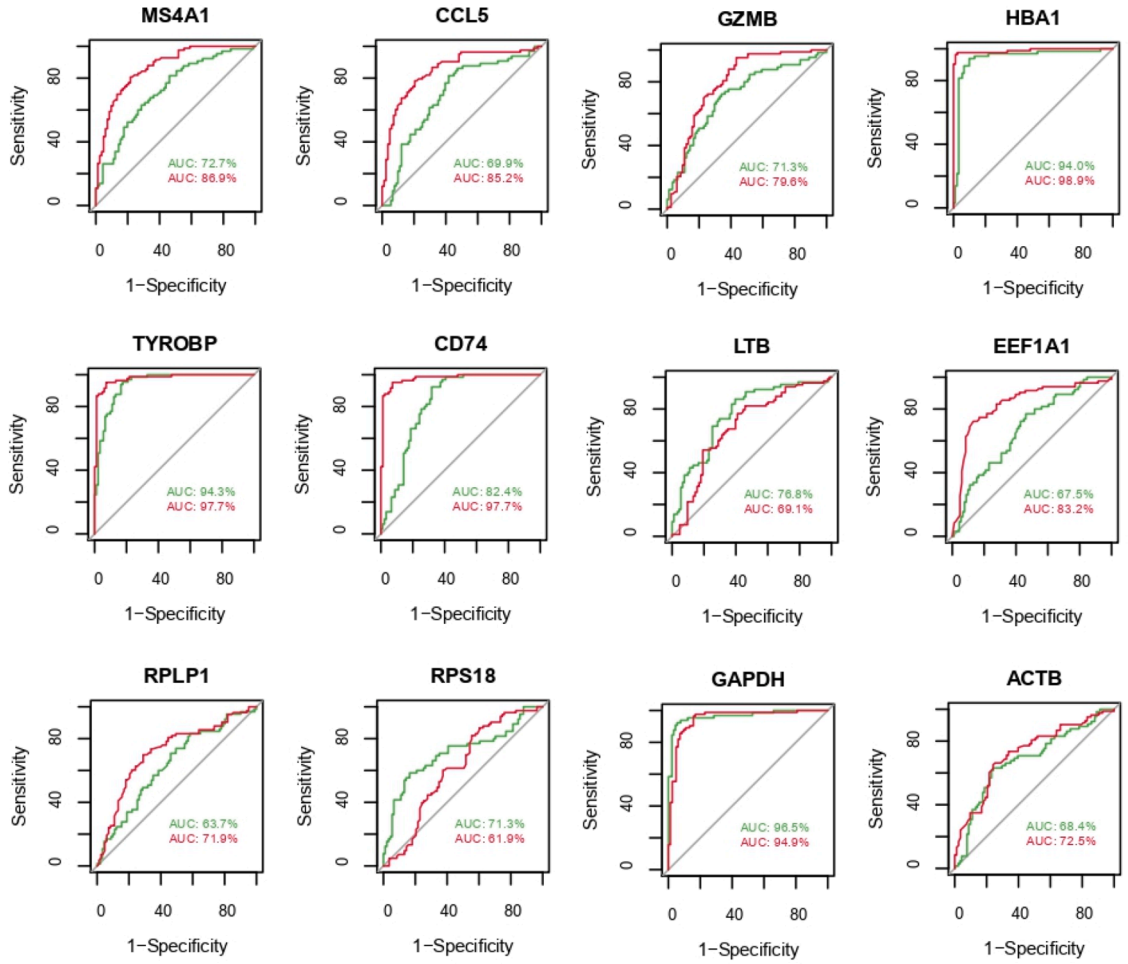

Fig. 5. ROC curve evaluating the diagnostic performance of the KGs in NSCLC. Red color indicates GSE75037 dataset and green color indicates GSE19188.

of cells. By comparing gene expression profiles, a total 158 DEGs containing 48 up- and 110 down-regulated genes were found. To infer the biological functions and pathways associated with NSCLC, GO and KEGG pathway enrichment analyses were performed. We identified 12 KGs (*MS4A1, CCL5, GZMB, HBA1, TYROBP, CD74, LTB, EEF1A1, RPLP1, RPS18, GAPDH,* and *ACTB*) based on the two methods of Betweenness and Stress using PPI network results, and the expression pattern and survival analysis of KGs were affirmed on the basis of the TCGA data. Using the TarBase and JASPAR databases in NetworkAnalyst, we identified 8 key miRNAs (hsa-miR-124-3p, hsa-miR-34a-5p, hsa-miR-21-5p, hsa-miR-155-5p, hsa-miR-449a, hsa-miR-24-3p, hsa-let-7b-5p, hsa-miR-7-5p) and 7 key TFs (FOXC1, YY1, CEBPB, TFAP2A, SREBF2, RELA, GATA2).

All the identified 12 KGs have been supported by different studies for

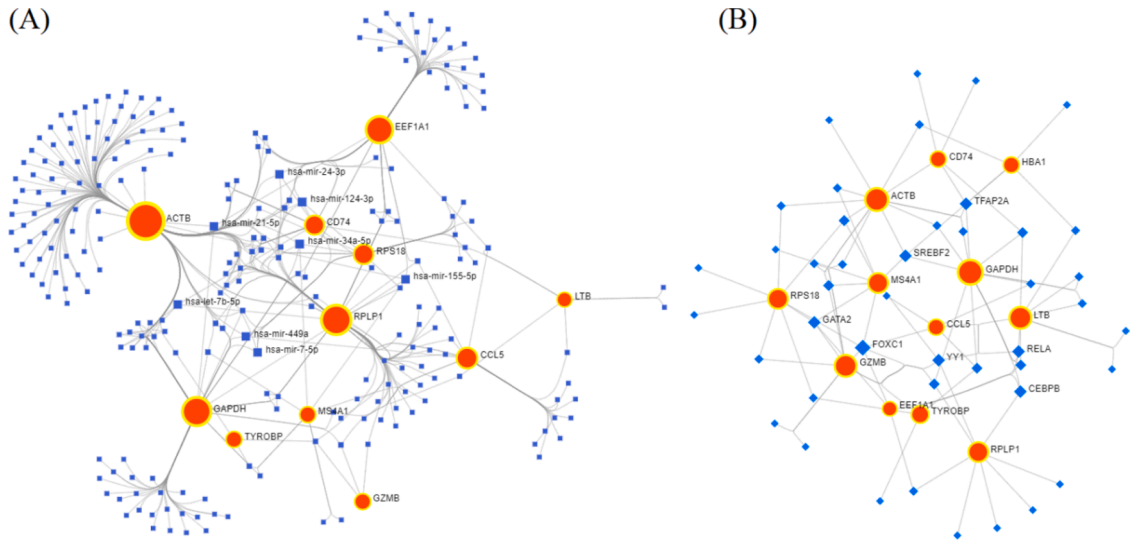

Fig. 6. Gene-miRNAs and Gene-TFs interaction network. (A) In the Gene-miRNAs interaction network, the red bigger circles indicate genes and the blue squares indicate miRNAs. (B) In Gene-TFs interaction network, the red circles indicate genes and the blue diamond shape indicates TFs.

Table 3
Targeted transcription factors (TFs) and miRNAs from regulatory interaction network (Gene-TFs and Gene-miRNAs).

S.N	TFs Label	Degree	S.N	miRNAs Label	Degree
1	FOXC1	6	1	hsa-miR-124-3p	5
2	YY1	4	2	hsa-miR-34a-5p	5
3	CEBPB	4	3	hsa-miR-21-5p	4
4	TFAP2A	4	4	hsa-miR-155-5p	4
5	SREBF2	4	5	hsa-miR-449a	4
6	RELA	4	6	hsa-miR-24-3p	4
7	GATA2	4	7	hsa-let-7b-5p	4
			8	hsa-miR-7-5p	4

lung and other cancers. The deregulation of *MS4A1* in lung squamous cell cancers might occur because of the expression of CD20 stromal lymphocytes [52]. Another study showed *MS4A1* as a prognostic biomarker for LUAD [53]. hsa-miR-147a can inhibit the outgrowth and metastasis of NSCLC by aiming the *CCL5* gene [54]. It has been found that *GZMB* is significantly associated with poor prognosis in SCLC [55]. Over expression of *HBA1* showed low overall survival in non-smoker female lung cancer patients [56]. *YAP1* promotes multidrug resistance of SCLC through signaling pathways associated with *CD74* [57]. *LTB-4* participates in the recruitment of neutrophils in the airways at NSCLC [58]. *ACTB, EEF1A1*, and *RPS18* are reported to be relevant genes for qRT-PCR analysis of lung cancer also *EEF1A1* is responsible for lung cancer development in smokers [59–61]. Aberrant methylation and high expression of *GAPDH* are associated with poor prognosis in LUAD patients [62]. *TYROBP* is a novel key gene with prognostic value in gastric cancer by integrated network analysis [63]. The gene biomarker *RPLP1* is an anti-metastasis candidate therapeutic target with poor prognosis in triple-negative breast cancer [64]. To our knowledge *TYROBP* and *RPLP1* are not yet been reported for lung cancer progression. Hence, we can say that these two genes (*TYROBP* and *RPLP1*) are novel with good prognostic value in our study for NSCLC.

Furthermore, hsa-miR-124-3p is reported as a tumor suppressor and inhibits the progression of several tumors, including NSCLC [65–67]. hsa-miR-34a-5p resists the brainstem glioma cell invasion [68]. hsa-miR-21-5p is one of the most important prognostic biomarkers for NSCLC [69]. hsa-miR-155-5p, hsa-miR-24-3p, and hsa-let-7a-5p were reported to be up-regulated in LUAD tissues [70]. hsa-miR-449a is reported as a genetic risk factor for gastric cancer [71]. hsa-miR-7-5p is a prognostic biomarker for small cell lung cancer [72]. Among the identified TFs, FOXC1 is one of the pioneer TF and plays an important role in the development of lung, breast, and prostate cancer [73]. The expression of FOXC1 is increased in NSCLC tissues, and it has an adverse relationship with survival [74]. TFAP2C is contributed to NSCLC tumorigenesis by downregulating numerous tumor silencers such as *GADD45B, PMAIP1*, and *XAF1* [75].

The GO functional enrichment and KEGG pathway analysis revealed that the KGs are related to Protein metabolism, MHC class I receptor activity, B cell receptor activity, viral mRNA Translation, 3′ -UTR-mediated translational regulation etc. pathways. Most of the DEGs (47.3%) have enriched with protein metabolism (associated with KGs: EEF1A1, RPS18, RPLP1, and GZMB) term, and several studies have further claimed that NSCLC-causing genes are enriched in protein metabolism term [76,77]. The prognostic effect of the KGs with TCGA datasets in LUAD showed the worst survival rate which indicates that these KGs might be the prognostic biomarkers in LUAD. The differential expression stated the discriminating power of the KGs. Finally, the diagnostic effects of KGs were assessed by ROC analysis. The AUC values in ROC analysis indicated a comparatively good prediction performance of the KGs in NSCLC patients with higher sensitivity and specificity. Therefore, our overall analysis will provide valuable insights into NSCLC progression, KGs, key miRNAs, key TFs might be a novel diagnostic and prognostic biomarkers as well as potential regulators for the progression, diagnosis, and prognosis of NSCLC. In this study, we predicted the results through computational analysis; hence, we cannot recommend for treatment directly. We emphasize for further assessed at the molecular level by the wet-lab experiments in prior to clinical investigation.

Conclusion

The scRNA-seq data of peripheral blood cell allows the identification of distinct cell types and provides a new perspective on the pathogenesis of NSCLC. On the viewpoint of clustering analysis, we conclude that NSCLC is heterogeneous in numerous aspects. Through bioinformatics analysis, we identified 12 KGs (*MS4A1, CCL5, GZMB, HBA1, TYROBP, CD74, LTB, EEF1A1, RPLP1, RPS18, GAPDH,* and *ACTB*); among them, there were 2 novel KGs (*TYROBP* and *RPLP1*). Their targeted miRNAs and TFs were also identified, which play a significant role in NSCLC. Survival and ROC analysis showed the prognostic and diagnostic effect of KGs. Our overall findings suggested that these KGs might be the prognostic and diagnostic biomarkers for NSCLC.

Ethical Approval and Consent to participate

Not Applicable

Consent for publication

All the authors have provided consent for the publication.

Availability of supporting data

The scRNA-seq data of NSCLC were obtained from the Gene Expression Omnibus (GEO) database (https://www.ncbi.nlm.nih.gov/geo/) with accession number GSE127471.

CRediT authorship contribution statement

Adiba Sultana: Data curation, Formal analysis, Writing – original draft, Writing – review & editing. **Md Shahin Alam:** Data curation, Writing – original draft, Writing – review & editing. **Xingyun Liu:** Formal analysis, Writing – review & editing. **Rohit Sharma:** Formal analysis, Writing – review & editing. **Rajeev K. Singla:** Formal analysis, Writing – original draft, Writing – review & editing. **Rohit Gundamaraju:** Formal analysis, Writing – review & editing. **Bairong Shen:** Conceptualization, Funding acquisition, Supervision, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors are thankful to the National Natural Science Foundation of China for proving funding support.

Funding

This work was supported by the National Natural Science Foundation of China (Grant No. 32070671 and 32270690).

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.tranon.2022.101571.

References

[1] J. Xu, H. Nie, J. He, X. Wang, K. Liao, L. Tu, et al., Using machine learning modeling to explore new immune-related prognostic markers in non-small cell lung cancer, Front. Oncol. 10 (2020), 550002, https://doi.org/10.3389/fonc.2020.550002. Epub 2020/11/21PubMed PMID33215029PubMed Central PMCIDPMCPMC7665579.

[2] F. Song, Z. Xuan, X. Yang, X. Ye, Z. Pan, Fang Q. Identification of key microRNAs and hub genes in non-small-cell lung cancer using integrative bioinformatics and functional analyses, J. Cell. Biochem. 121 (3) (2020) 2690–2703, https://doi.org/10.1002/jcb.29489. Epub 2019/11/07PubMed PMID31692035.

[3] Y. Wang, L. Huang, S. Wu, Y. Jia, Y. Yang, L. Luo, et al., Bioinformatics analyses of the role of vascular endothelial growth factor in patients with non-small cell lung cancer, PLoS One 10 (9) (2015), e0139285, https://doi.org/10.1371/journal.pone.0139285. Epub 2015/10/01PubMed PMID26422603PubMed Central PMCIDPMCPMC4589385.

[4] Q. Sun, X. Li, M. Xu, L. Zhang, H. Zuo, Y. Xin, et al., Differential expression and bioinformatics analysis of circRNA in non-small cell lung cancer, Front. Genet. 11 (2020), 586814, https://doi.org/10.3389/fgene.2020.586814. Epub 2020/12/18PubMed PMID33329727PubMed Central PMCIDPMCPMC7732606.

[5] Q. Wu, B. Zhang, Y. Sun, R. Xu, X. Hu, S. Ren, et al., Identification of novel biomarkers and candidate small molecule drugs in non-small-cell lung cancer by integrated microarray analysis, Onco Targets Ther. 12 (2019) 3545–3563, https://doi.org/10.2147/OTT.S198621. Epub 2019/06/14PubMed PMID31190860PubMed Central PMCIDPMCPMC6526173.

[6] Xi Chen, Sarah A. Teichmann, K.B. Meyer, From tissues to cell types and back: single-cell gene expression analysis of tissue architecture, Annu. Rev. Biomed. Data Sci. 1 (2018) 29–51, https://doi.org/10.1146/annurev-biodatasci-080917-013452.

[7] T.K. Olsen, N. Baryawno, Introduction to single-cell RNA sequencing, Curr. Protoc. Mol. Biol. 122 (1) (2018) e57, https://doi.org/10.1002/cpmb.57. Epub 2018/06/01PubMed PMID29851283.

[8] C. Cochain, E. Vafadarnejad, P. Arampatzi, J. Pelisek, H. Winkels, K. Ley, et al., Single-Cell RNA-Seq reveals the transcriptional landscape and heterogeneity of aortic macrophages in murine atherosclerosis, Circ. Res. 122 (12) (2018) 1661–1674, https://doi.org/10.1161/CIRCRESAHA.117.312509. Epub 2018/03/17PubMed PMID29545365.

[9] J.W. Min, W.J. Kim, J.A. Han, Y.J. Jung, K.T. Kim, W.Y. Park, et al., Identification of distinct tumor subpopulations in lung adenocarcinoma via single-cell RNA-seq, PLoS One 10 (8) (2015), e0135817, https://doi.org/10.1371/journal.pone.0135817. Epub 2015/08/26PubMed PMID26305796PubMed Central PMCIDPMCPMC4549254.

[10] A.P. Patel, I. Tirosh, J.J. Trombetta, A.K. Shalek, S.M. Gillespie, H. Wakimoto, et al., Single-Cell RNA-Seq highlights intratumoral heterogeneity in primary glioblastoma, Science 344 (6190) (2014) 1396–1401, https://doi.org/10.1126/science.1254257. Epub 2014/06/14PubMed PMID24925914PubMed Central PMCIDPMCPMC4123637.

[11] K. Gong, H. Zhou, H. Liu, T. Xie, Y. Luo, H. Guo, et al., Identification and integrate analysis of key biomarkers for diagnosis and prognosis of non-small cell lung cancer based on bioinformatics analysis, Technol. Cancer Res. Treat. 20 (2021), 15330338211060202, https://doi.org/10.1177/15330338211060202. Epub 2021/11/27PubMed PMID34825846PubMed Central PMCIDPMCPMC8649439.

[12] J. Kim, Z. Xu, P.A. Marignani, Single-cell RNA sequencing for the identification of early-stage lung cancer biomarkers from circulating blood, NPJ Genom. Med. 6 (1) (2021) 87, https://doi.org/10.1038/s41525-021-00248-y. Epub 2021/10/17PubMed PMID34654834PubMed Central PMCIDPMCPMC8519939.

[13] Z. Chen, C.M. Fillmore, P.S. Hammerman, C.F. Kim, K.K. Wong, Non-small-cell lung cancers: a heterogeneous set of diseases, Nat. Rev. Cancer 14 (8) (2014) 535–546, https://doi.org/10.1038/nrc3775. Epub 2014/07/25PubMed PMID25056707PubMed Central PMCIDPMCPMC5712844.

[14] G.W. Warren, K.M. Cummings, Tobacco and lung cancer: risks, trends, and outcomes in patients with cancer, Am. Soc. Clin. Oncol. Educ. Book 33 (2013) 359–364, https://doi.org/10.14694/EdBook_AM.2013.33.359.

[15] B.A. Chan, B.G. Hughes, Targeted therapy for non-small cell lung cancer: current standards and the promise of the future, Transl. Lung Cancer Res. 4 (1) (2015) 36–54, https://doi.org/10.3978/j.issn.2218-6751.2014.05.01. Epub 2015/03/26PubMed PMID25806345PubMed Central PMCIDPMCPMC4367711.

[16] F.R. Hirsch, A. Spreafico, S. Novello, M.D. Wood, L. Simms, M. Papotti, The prognostic and predictive role of histology in advanced non-small cell lung cancer: a literature review, J. Thorac. Oncol. 3 (12) (2008) 1468–1481, https://doi.org/10.1097/JTO.0b013e318189f551. Epub 2008/12/06PubMed PMID19057275.

[17] D.S. Ettinger, W. Akerley, H. Borghaei, A.C. Chang, R.T. Cheney, L.R. Chirieac, et al., Non-small cell lung cancer, J. Natl. Compr. Cancer Netw. 10 (10) (2012) 1236–1271, https://doi.org/10.6004/jnccn.2012.0130. Epub 2012/10/12PubMed PMID23054877.

[18] American Cancer Society. Lung Cancer Survival Rates 2019. Available from: https://www.cancer.org/cancer/lung-cancer/detection-diagnosis-staging/survival-rates.html.

[19] J. Hayes, P.P. Peruzzi, S. Lawler, MicroRNAs in cancer: biomarkers, functions and therapy, Trends Mol. Med. 20 (8) (2014) 460–469, https://doi.org/10.1016/j.molmed.2014.06.005. Epub 2014/07/17PubMed PMID25027972.

[20] W.C. Cho, OncomiRs: the discovery and progress of microRNAs in cancers, Mol. Cancer 6 (2007) 60, https://doi.org/10.1186/1476-4598-6-60. Epub 2007/09/27PubMed PMID17894887PubMed Central PMCIDPMCPMC2098778.

[21] A. Esquela-Kerscher, F.J. Slack, Oncomirs - microRNAs with a role in cancer, Nat. Rev. Cancer 6 (4) (2006) 259–269, https://doi.org/10.1038/nrc1840. Epub 2006/03/25PubMed PMID16557279.

[22] Y. Karube, H. Tanaka, H. Osada, S. Tomida, Y. Tatematsu, K. Yanagisawa, et al., Reduced expression of Dicer associated with poor prognosis in lung cancer patients, Cancer Sci. 96 (2) (2005) 111–115, https://doi.org/10.1111/j.1349-7006.2005.00015.x. Epub 2005/02/23PubMed PMID15723655.

[23] L. Du, J.J. Schageman, Girard L. Irnov, S.M. Hammond, J.D. Minna, et al., MicroRNA expression distinguishes SCLC from NSCLC lung tumor cells and suggests a possible pathological relationship between SCLCs and NSCLCs, J. Exp. Clin. Cancer Res. 29 (2010) 75, https://doi.org/10.1186/1756-9966-29-75. Epub 2010/07/14PubMed PMID20624269PubMed Central PMCIDPMCPMC2907339.

[24] M.S. Alam, M.M. Rahaman, A. Sultana, G. Wang, M.N.H. Mollah, Statistics and network-based approaches to identify molecular mechanisms that drive the progression of breast cancer, Comput. Biol. Med. 145 (2022), 105508, https://doi.org/10.1016/j.compbiomed.2022.105508. Epub 2022/04/22PubMed PMID35447458.

[25] M.S. Alam, A. Sultana, M.S. Reza, M. Amanullah, S.R. Kabir, M.N.H. Mollah, Integrated bioinformatics and statistical approaches to explore molecular biomarkers for breast cancer diagnosis, prognosis and therapies, PLoS One 17 (5) (2022), e0268967, https://doi.org/10.1371/journal.pone.0268967. Epub 2022/05/27PubMed PMID35617355.

[26] W. Chen, S. Zhu, Y. Zhang, J. Xiao, D. Tian, Identification of key candidate tumor biomarkers in non-small-cell lung cancer by in silico analysis, Oncol. Lett. 19 (1) (2020) 1008–1016, https://doi.org/10.3892/ol.2019.11169. Epub 2020/01/04PubMed PMID31897214PubMed Central PMCIDPMCPMC6924182.

[27] M. Yang, H. Sun, J. He, H. Wang, X. Yu, L. Ma, et al., Interaction of ribosomal protein L22 with casein kinase 2alpha: a novel mechanism for understanding the biology of non-small cell lung cancer, Oncol. Rep. 32 (1) (2014) 139–144, https://doi.org/10.3892/or.2014.3187. Epub 2014/05/21PubMed PMID24840952.

[28] Z. Yang, N. Yang, Q. Ou, Y. Xiang, T. Jiang, X. Wu, et al., Investigating novel resistance mechanisms to third-generation EGFR tyrosine kinase inhibitor Osimertinib in non-small cell lung cancer patients, Clin. Cancer Res. 24 (13) (2018) 3097–3107, https://doi.org/10.1158/1078-0432.CCR-17-2310. Epub 2018/03/07PubMed PMID29506987.

[29] X. Zhang, D. Zhang, L. Huang, G. Li, L. Chen, J. Ma, et al., Discovery of novel biomarkers of therapeutic responses in Han Chinese pemetrexed-based treated advanced NSCLC patients, Front. Pharmacol. 10 (2019) 944, https://doi.org/10.3389/fphar.2019.00944. Epub 2019/09/12PubMed PMID31507426PubMed Central PMCIDPMCPMC6716463.

[30] Y. Shen, X. Pan, J. Yang, Gene regulation and prognostic indicators of lung squamous cell carcinoma: TCGA-derived miRNA/mRNA sequencing and DNA methylation data, J. Cell. Physiol. 234 (12) (2019) 22896–22910, https://doi.org/10.1002/jcp.28852. Epub 2019/06/07PubMed PMID31169310.

[31] K. Valk, T. Vooder, R. Kolde, M.A. Reintam, C. Petzold, J. Vilo, et al., Gene expression profiles of non-small cell lung cancer: survival prediction and new biomarkers, Oncology 79 (3-4) (2010) 283–292, https://doi.org/10.1159/000322116. Epub 2011/03/18PubMed PMID21412013.

[32] R. Puzone, G. Savarino, S. Salvi, M.G. Dal Bello, G. Barletta, C. Genova, et al., Glyceraldehyde-3-phosphate dehydrogenase gene over expression correlates with poor prognosis in non small cell lung cancer patients, Mol. Cancer 12 (1) (2013) 97, https://doi.org/10.1186/1476-4598-12-97. Epub 2013/08/31PubMed PMID23988223PubMed Central PMCIDPMCPMC3766010.

[33] Newman AM, Steen CB, Liu CL, Gentles AJ, Chaudhuri AA, Scherer F, et al. High-throughput tissue dissection and cell purification with digital cytometry [scRNA-Seq] 2019. Available from: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE127471.

[34] A. Butler, P. Hoffman, P. Smibert, E. Papalexi, R. Satija, Integrating single-cell transcriptomic data across different conditions, technologies, and species, Nat. Biotechnol. 36 (5) (2018) 411–420, https://doi.org/10.1038/nbt.4096. Epub 2018/04/03PubMed PMID: 29608179PubMed Central PMCIDPMCPMC6700744.

[35] X. Qiu, Q. Mao, Y. Tang, L. Wang, R. Chawla, H.A. Pliner, et al., Reversed graph embedding resolves complex single-cell trajectories, Nat. Methods 14 (10) (2017) 979–982, https://doi.org/10.1038/nmeth.4402. Epub 2017/08/22PubMed PMID28825705PubMed Central PMCIDPMCPMC5764547.

[36] M. Pathan, S. Keerthikumar, C.S. Ang, L. Gangoda, C.Y. Quek, N.A. Williamson, et al., FunRich: an open access standalone functional enrichment and interaction network analysis tool, Proteomics 15 (15) (2015) 2597–2601, https://doi.org/10.1002/pmic.201400515. Epub 2015/04/30PubMed PMID25921073.

[37] C. Gene Ontology, The Gene Ontology (GO) project in 2006, Nucleic Acids Res. 34 (2006) D322, https://doi.org/10.1093/nar/gkj021. Database issue-6Epub 2005/12/31PubMed PMID16381878PubMed Central PMCIDPMCPMC1347384.

[38] M. Kanehisa, S. Goto, KEGG: Kyoto encyclopedia of genes and genomes, Nucleic Acids Res. 28 (1) (2000) 27–30, https://doi.org/10.1093/nar/28.1.27. Epub 1999/12/11PubMed PMID10592173PubMed Central PMCIDPMCPMC102409.

[39] D. Szklarczyk, A.L. Gable, D. Lyon, A. Junge, S. Wyder, J. Huerta-Cepas, et al., STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets, Nucleic Acids Res. 47 (D1) (2019) D607–DD13, https://doi.org/10.1093/nar/gky1131. Epub 2018/11/27PubMed PMID30476243PubMed Central PMCIDPMCPMC6323986.

[40] P. Shannon, A. Markiel, O. Ozier, N.S. Baliga, J.T. Wang, D. Ramage, et al., Cytoscape: a software environment for integrated models of biomolecular interaction networks, Genome Res. 13 (11) (2003) 2498–2504, https://doi.org/10.1101/gr.1239303. Epub 2003/11/05PubMed PMID14597658PubMed Central PMCIDPMCPMC403769.

[41] C.H. Chin, S.H. Chen, H.H. Wu, C.W. Ho, M.T. Ko, C.Y. Lin, cytoHubba: identifying hub objects and sub-networks from complex interactome, BMC Syst. Biol. 8 (4) (2014) S11, https://doi.org/10.1186/1752-0509-8-S4-S11. SupplEpub 2014/12/19PubMed PMID25521941PubMed Central PMCIDPMCPMC4290687.

[42] R. Aguirre-Gamboa, H. Gomez-Rueda, E. Martinez-Ledesma, A. Martinez-Torteya, R. Chacolla-Huaringa, A. Rodriguez-Barrientos, et al., SurvExpress: an online biomarker validation tool and database for cancer gene expression data using survival analysis, PLoS One 8 (9) (2013) e74250, https://doi.org/10.1371/journal.pone.0074250. Epub 2013/09/26PubMed PMID24066126PubMed Central PMCIDPMCPMC3774754.

[43] X. Robin, N. Turck, A. Hainard, N. Tiberti, F. Lisacek, J.C. Sanchez, et al., pROC: an open-source package for R and S+ to analyze and compare ROC curves, BMC Bioinform. 12 (2011) 77, https://doi.org/10.1186/1471-2105-12-77. Epub 2011/03/19PubMed PMID21414208PubMed Central PMCIDPMCPMC3068975.

[44] G. Zhou, O. Soufan, J. Ewald, R.E.W Hancock, N. Basu, J. Xia, NetworkAnalyst 3.0: a visual analytics platform for comprehensive gene expression profiling and meta-analysis, Nucleic Acids Res. 47 (W1) (2019) W234–WW41, https://doi.org/10.1093/nar/gkz240. Epub 2019/04/02PubMed PMID30931480PubMed Central PMCIDPMCPMC6602507.

[45] D. Karagkouni, M.D. Paraskevopoulou, S. Chatzopoulos, I.S. Vlachos, S. Tastsoglou, I. Kanellos, et al., DIANA-TarBase v8: a decade-long collection of experimentally supported miRNA-gene interactions, Nucleic Acids Res. 46 (D1) (2018) D239–DD45, https://doi.org/10.1093/nar/gkx1141. Epub 2017/11/21PubMed PMID29156006PubMed Central PMCIDPMCPMC5753203.

[46] S.D. Hsu, F.M. Lin, W.Y. Wu, C. Liang, W.C. Huang, W.L. Chan, et al., miRTarBase: a database curates experimentally validated microRNA-target interactions, Nucleic Acids Res. 39 (2011) D163–D169, https://doi.org/10.1093/nar/gkq1107. Database issueEpub 2010/11/13PubMed PMID21071411PubMed Central PMCIDPMCPMC3013699.

[47] F. Xiao, Z. Zuo, G. Cai, S. Kang, X. Gao, T. Li, miRecords: an integrated resource for microRNA-target interactions, Nucleic Acids Res. 37 (2009) D105–D110, https://doi.org/10.1093/nar/gkn851. Database issueEpub 2008/11/11PubMed PMID18996891PubMed Central PMCIDPMCPMC2686554.

[48] E.P. Consortium, The ENCODE (ENCyclopedia Of DNA Elements) project, Science 306 (5696) (2004) 636–640, https://doi.org/10.1126/science.1105136. Epub 2004/10/23PubMed PMID15499007.

[49] O. Fornes, J.A. Castro-Mondragon, A. Khan, R. van der Lee, X. Zhang, P. A. Richmond, et al., JASPAR 2020: update of the open-access database of transcription factor binding profiles, Nucleic Acids Res. 48 (D1) (2020) D87–D92, https://doi.org/10.1093/nar/gkz1001. Epub 2019/11/09PubMed PMID31701148PubMed Central PMCIDPMCPMC7145627.

[50] A. Lachmann, H. Xu, J. Krishnan, S.I. Berger, A.R. Mazloom, Ma'ayan A. ChEA, transcription factor regulation inferred from integrating genome-wide ChIP-X experiments, Bioinformatics 26 (19) (2010) 2438–2444, https://doi.org/10.1093/bioinformatics/btq466. Epub 2010/08/17PubMed PMID20709693PubMed Central PMCIDPMCPMC2944209.

[51] A.L. Liang, S.L. Du, B. Zhang, J. Zhang, X. Ma, C.Y. Wu, et al., Screening miRNAs associated with resistance gemcitabine from exosomes in A549 lung cancer cells, Cancer Manag. Res. 11 (2019) 6311–6321, https://doi.org/10.2147/CMAR.S209149. Epub 2019/08/03PubMed PMID31372037PubMed Central PMCIDPMCPMC6626902.

[52] C.M. Wright, S.M. Savarimuthu Francis, M.E. Tan, M.U. Martins, C. Winterford, M. R. Davidson, et al., MS4A1 dysregulation in asbestos-related lung squamous cell carcinoma is due to CD20 stromal lymphocyte expression, PLoS One 7 (4) (2012) e34943, https://doi.org/10.1371/journal.pone.0034943. Epub 2012/04/20PubMed PMID22514692PubMed Central PMCIDPMCPMC3325913.

[53] C. Ma, H. Luo, J. Cao, X. Zheng, J. Zhang, Y. Zhang, et al., Identification of a novel tumor microenvironment-associated eight-gene signature for prognosis prediction in lung adenocarcinoma, Front. Mol. Biosci. 7 (2020), 571641, https://doi.org/10.3389/fmolb.2020.571641. Epub 2020/10/27PubMed PMID33102522PubMed Central PMCIDPMCPMC7546815.

[54] Y. Lu, X.R. Luan, miR-147a suppresses the metastasis of non-small-cell lung cancer by targeting CCL5, J. Int. Med. Res. 48 (4) (2020), 300060519883098, https://doi.org/10.1177/0300060519883098. Epub 2019/12/31PubMed PMID31884861PubMed Central PMCIDPMCPMC7607764.

[55] Y. Song, Y. Sun, T. Sun, R. Tang, Comprehensive bioinformatics analysis identifies tumor microenvironment and immune-related genes in small cell lung cancer, Comb. Chem. High Throughput Screen. 23 (5) (2020) 381–391, https://doi.org/10.2174/1386207323666200407075004. Epub 2020/04/09PubMed PMID32264809.

[56] K. Shi, N. Li, M. Yang, W. Li, Identification of key genes and pathways in female lung cancer patients who never smoked by a bioinformatics analysis, J. Cancer 10 (1) (2019) 51–60, https://doi.org/10.7150/jca.26908. Epub 2019/01/22PubMed PMID30662525PubMed Central PMCIDPMCPMC6329865.

[57] Y. Song, Y. Sun, Y. Lei, K. Yang, R. Tang, YAP1 promotes multidrug resistance of small cell lung cancer by CD74-related signaling pathways, Cancer Med. 9 (1) (2020) 259–268, https://doi.org/10.1002/cam4.2668. Epub 2019/11/07PubMed PMID31692299PubMed Central PMCIDPMCPMC6943160.

[58] G.E. Carpagnano, G.P. Palladino, D. Lacedonia, A. Koutelou, S. Orlando, M. P. Foschino-Barbaro, Neutrophilic airways inflammation in lung cancer: the role of exhaled LTB-4 and IL-8, BMC Cancer 11 (2011) 226, https://doi.org/10.1186/1471-2407-11-226. Epub 2011/06/09PubMed PMID21649887PubMed Central PMCIDPMCPMC3130703.

[59] C. Zhan, Y. Zhang, J. Ma, L. Wang, W. Jiang, Y. Shi, et al., Identification of reference genes for qRT-PCR in human lung squamous-cell carcinoma by RNA-Seq, Acta Biochim. Biophys. Sin. 46 (4) (2014) 330–337, https://doi.org/10.1093/abbs/gmt153. Epub 2014/01/25PubMed PMID24457517.

[60] R. Chari, K.M. Lonergan, L.A. Pikor, B.P. Coe, C.Q. Zhu, T.H. Chan, et al., A sequence-based approach to identify reference genes for gene expression analysis, BMC Med. Genom. 3 (2010) 32, https://doi.org/10.1186/1755-8794-3-32. Epub 2010/08/05PubMed PMID20682026PubMed Central PMCIDPMCPMC2928167.

[61] Z. Yang, B. Zhuan, Y. Yan, S. Jiang, T. Wang, Identification of gene markers in the development of smoking-induced lung cancer, Gene 576 (1) (2016) 451–457, https://doi.org/10.1016/j.gene.2015.10.060. Pt 3Epub 2015/11/01PubMed PMID26518718.

[62] X. Wang, D. Shi, D. Zhao, D. Hu, Aberrant methylation and differential expression of SLC2A1, TNS4, GAPDH, ATP8A2, and CASZ1 are associated with the prognosis of lung adenocarcinoma, Biomed. Res. Int. 2020 (2020), 1807089, https://doi.org/10.1155/2020/1807089. Epub 2020/10/09PubMed PMID33029490PubMed Central PMCIDPMCPMC7532994.

[63] J. Jiang, Y. Ding, M. Wu, X. Lyu, H. Wang, Y. Chen, et al., Identification of TYROBP and C1QB as two novel key genes with prognostic value in gastric cancer by network analysis, Front. Oncol. 10 (2020) 1765, https://doi.org/10.3389/fonc.2020.01765. Epub 2020/10/06PubMed PMID33014868PubMed Central PMCIDPMCPMC7516284.

[64] Z. He, Q. Xu, X. Wang, J. Wang, X. Mu, Y. Cai, et al., RPLP1 promotes tumor metastasis and is associated with a poor prognosis in triple-negative breast cancer patients, Cancer Cell Int. 18 (2018) 170, https://doi.org/10.1186/s12935-018-0658-0. Epub 2018/11/06PubMed PMID30386179PubMed Central PMCIDPMCPMC6203216.

[65] R.Q. He, X. Yang, L. Liang, G. Chen, J. Ma, MicroRNA-124-3p expression and its prospective functional pathways in hepatocellular carcinoma: a quantitative polymerase chain reaction, gene expression omnibus and bioinformatics study, Oncol. Lett. 15 (4) (2018) 5517–5532, https://doi.org/10.3892/ol.2018.8045. Epub 2018/03/20PubMed PMID29552191PubMed Central PMCIDPMCPMC5840674.

[66] Y. Sun, X. Ai, S. Shen, S. Lu, NF-kappaB-mediated miR-124 suppresses metastasis of non-small-cell lung cancer by targeting MYO10, Oncotarget 6 (10) (2015) 8244–8254, https://doi.org/10.18632/oncotarget.3135. Epub 2015/03/10PubMed PMID25749519PubMed Central PMCIDPMCPMC4480748.

[67] L.X. Tang, G.H. Chen, H. Li, P. He, Y. Zhang, X.W. Xu, Long non-coding RNA OGFRP1 regulates LYPD3 expression by sponging miR-124-3p and promotes non-small cell lung cancer progression, Biochem. Biophys. Res. Commun. 505 (2) (2018) 578–585, https://doi.org/10.1016/j.bbrc.2018.09.146. Epub 2018/10/03PubMed PMID30274775.

[68] X. Chen, D. Dong, C. Pan, C. Xu, Y. Sun, Y. Geng, et al., Identification of grade-associated MicroRNAs in brainstem gliomas based on microarray data, J. Cancer 9 (23) (2018) 4463–4476, https://doi.org/10.7150/jca.26417. Epub 2018/12/07PubMed PMID30519352PubMed Central PMCIDPMCPMC6277643.

[69] K. Wang, M. Chen, W. Wu, Analysis of microRNA (miRNA) expression profiles reveals 11 key biomarkers associated with non-small cell lung cancer, World J. Surg. Oncol. 15 (1) (2017) 175, https://doi.org/10.1186/s12957-017-1244-y. Epub 2017/09/21PubMed PMID28927412PubMed Central PMCIDPMCPMC5606074.

[70] A. Zeybek, N. Oz, S. Kalemci, T. Edgunlu, M.T. Kiziltug, K. Tosun, et al., Diagnostic value of MiR-125b as a potential biomarker for stage I lung adenocarcinoma, Curr. Mol. Med. 19 (3) (2019) 216–227, https://doi.org/10.2174/1566524019666190314113800. Epub 2019/03/15PubMed PMID30868951.

[71] J. Shi, Y. Liu, J. Liu, J. Zhou, Hsa-miR-449a genetic variant is associated with risk of gastric cancer in a Chinese population, Int. J. Clin. Exp. Pathol. 8 (10) (2015) 13387–13392. Epub 2016/01/02PubMed PMID26722545PubMed Central PMCIDPMCPMC4680490.

[72] X. Li, C. Ma, H. Luo, J. Zhang, J. Wang, H. Guo, Identification of the differential expression of genes and upstream microRNAs in small cell lung cancer compared with normal lung based on bioinformatics analysis, Medicine (Baltimore). 99 (11) (2020) e19086, https://doi.org/10.1097/MD.0000000000019086. Epub 2020/03/17PubMed PMID32176034PubMed Central PMCIDPMCPMC7440067.

[73] Y. Zhang, Y.X. Huang, D.L. Wang, B. Yang, H.Y. Yan, L.H. Lin, et al., LncRNA DSCAM-AS1 interacts with YBX1 to promote cancer progression by forming a positive feedback loop that activates FOXA1 transcription network, Theranostics 10 (23) (2020) 10823–10837, https://doi.org/10.7150/thno.47830. Epub 2020/09/16PubMed PMID32929382PubMed Central PMCIDPMCPMC7482804.

[74] S. Cao, Z. Wang, X. Gao, W. He, Y. Cai, H. Chen, et al., FOXC1 induces cancer stem cell-like properties through upregulation of beta-catenin in NSCLC, J. Exp. Clin. Cancer Res. 37 (1) (2018) 220, https://doi.org/10.1186/s13046-018-0894-0. Epub 2018/09/08PubMed PMID30189871PubMed Central PMCIDPMCPMC6127900.

[75] H. Do, D. Kim, J. Kang, B. Son, D. Seo, H. Youn, et al., TFAP2C increases cell proliferation by downregulating GADD45B and PMAIP1 in non-small cell lung cancer cells, Biol. Res. 52 (1) (2019) 35, https://doi.org/10.1186/s40659-019-0244-5. Epub 2019/07/13PubMed PMID31296259PubMed Central PMCIDPMCPMC6625030.

[76] Q. Zhang, J. Liu, R. Li, R. Zhao, M. Zhang, S. Wei, et al., A network pharmacology approach to investigate the anticancer mechanism and potential active ingredients of Rheum palmatum L. against lung cancer via induction of apoptosis, Front. Pharmacol. 11 (2020), 528308, https://doi.org/10.3389/fphar.2020.528308. Epub 2020/12/01PubMed PMID33250766PubMed Central PMCIDPMCPMC7672213.

[77] B.Y. Kim, J. Lee, N.S. Kim, Helveticoside is a biologically active component of the seed extract of Descurainia sophia and induces reciprocal gene regulation in A549 human lung cancer cells, BMC Genom. 16 (2015) 713, https://doi.org/10.1186/s12864-015-1918-1. Epub 2015/09/20PubMed PMID26384484PubMed Central PMCIDPMCPMC4575430.