Cumhuriyet Science Journal

e-ISSN: 2587-246X
ISSN: 2587-2680

Cumhuriyet Sci. J., 42(4) (2021) 751-757
http://dx.doi.org/10.17776/csj.887810

Identification and validation of key genes associated with smoking-induced lung adenocarcinoma development through bioinformatics analysis and predictions of small-molecule drugs

Hamid CEYLAN [1,*]

[1]*Atatürk University, Faculty of Science, Department of Molecular Biology and Genetics, 25240, Erzurum, TURKEY*

Abstract

Although smoking is known to be the leading risk factor for lung cancer, it is still unclear how normal cells turn cancerous in cigarette smokers. This study aimed to identify key molecular drivers that contributed to the progression and prognosis of lung adenocarcinoma (LUAD) in cigarette smokers, as well as screen, correlated small molecule therapeutic drugs by bioinformatics analysis. Gene expression profile was obtained from the Gene Expression Omnibus (GEO) database. Differentially expressed genes (DEGs) between current smokers without cancer and never smokers were identified and were analyzed to identify gene ontologies, pathways, protein-protein interaction (PPI) networks, hub genes, and prognostic potentials. Finally, effective small-molecule compounds were screened by the Connectivity Map (CMap) database. A total of nine genes were screened out as the critical among the DEGs from the PPI network. Overall survival analysis revealed that high mRNA expression of ACTR2 and ANAPC10 were significantly associated with the LUAD. Furthermore, three candidate small-molecule drugs for manipulating LUAD progression were predicted. Identification of critical genes involved in disease development and candidate drugs to combat it can lead us to better diagnosis and targeted therapy strategies. The results of the present study may provide insight into the mechanisms underlying LUAD pathogenesis development risk in cigarette smokers and may provide potential targets for prevention.

Article info

History:
Received: 27.02.2021
Accepted: 02.10.2021

Keywords:
Bioinformatics analysis,
Differentially expressed genes,
Lung adenocarcinoma,
Overall survival,
Smoking.

1. Introduction

Lung cancer is the second most frequently diagnosed cancer type in 2020 (2.2 million new cases; 11.4%) and the common reason for cancer-related deaths (1.8 million deaths; 18%) [1]. About 85% of patients were diagnosed with the non-small cell lung cancer (NSCLC) subtype. Lung adenocarcinoma (LUAD), a subtype of NSCLC, is responsible for about 40% of all lung cancer cases [2]. For this reason, determining the molecular drivers and mechanisms underlying cancer progression in LUAD is considered very important in terms of diagnosis and developing effective treatment approaches.

Smoking is an extremely important risk factor, which causes approximately 80-90% of lung cancer cases worldwide and is the highest source of cancer-related deaths in humans [3]. Individuals with a history of smoking are tens of times more likely to develop LC than never smokers. Lung cancer-related genomic alterations have a distinct difference between smokers and non-smokers [4]. However, information on the underlying molecular mechanisms that contribute to lung tumor formation in smokers is still quite insufficient.

Polygenic or multifactorial diseases, including cancer, arises as a result of complex interactions of multiple genes [5]. Given the complexity observed in cancer pathophysiology, alterations in the whole genome and transcriptome should be taken into account for effective treatment [6]. Identifying differentially expressed genes by analyzing DNA microarray datasets using bioinformatics tools is considered an approach that has the potential to reveal specific mechanisms and molecular events precisely in terms of disease management [7, 8]. This approach can provide a strong framework for understanding how a pathological process is regulated, as well as identifying

*Corresponding author. *e-mail address: hamid.ceylan@atauni.edu.tr*

biomarkers that can be used in diagnosis, and identifying potential therapeutic targets and tools.

In this study, it was aimed to investigate the global gene expression differences in bronchial epithelial samples of current smokers and never smokers to elucidate the biological mechanisms underlying lung adenocarcinoma that may be caused by smoking. The selected microarray dataset was analyzed and the DEGs were identified. A PPI network was created to elucidate important relationships between DEGs and identify key genes, and also the expression and prognostic potential of key genes were studied. Finally, candidate small molecules that can be used in the prevention of smoking-induced LUAD development were predicted.

2. Methods

2.1. Microarray data processing and identification of DEGs

The gene expression profile (GSE19027) [9] was downloaded from the NCBI GEO (http://www.ncbi.nlm.nih.gov/geo) public repository. In this study, to detect genes whose expression is altered only by smoking, datasets from never-smokers and datasets from patients who smoke but do not have cancer were selected. A total of 25 bronchial epithelial tissue samples (6 never smokers and 19 current smokers without cancer) were used for analysis. Identification of DEGs between smokers and non-smokers was performed using the GEO2R web tool (https://www.ncbi.nlm.nih.gov/geo/geo2r/). The cut-off criteria were set as follows $|\text{LogFC}| > 1.5$ and $p<0.05$.

2.2. PPI network construction and module analysis

STRING (Search Tool for the Retrieval of Interacting Genes; https://string-db.org/) database was employed to evaluate the interrelationships between DEGs. Cytoscape software was used to analyze and visualize the PPI network. Finally, Molecular Complex Detection (MCODE) plugin of Cytoscape was used to filter central modules in the PPI network.

2.3. Hub gene selection and analysis

CytoHubba plugin of Cytoscape was used to identify the hub genes. The Database for Annotation, Visualization and Integrated Discovery database (DAVID; https://david.ncifcrf.gov/home.jsp) application was used to perform GO (gene ontology) and Kyoto Encyclopedia of Genes and Genome (KEGG) pathway enrichment analysis of hub genes.

2.4. Survival analysis and validation of hub genes

To confirm the reliability of the hub genes, their expressions were validated using GEPIA (Gene Expression Profiling Interactive Analysis) database [10]. In addition, GEPIA and Kaplan-Meier [11] survival curves of overall survival were used to analyze prognostic potentials and survival differences of the hub genes.

2.5. Candidate small molecule drugs prediction

The CMAP online tool (https://www.broadinstitute.org/connectivity-map-cmap) which contains whole genomic expression profiles for small active molecular inferences, was used to mine potential small molecules. Hub genes probesets in GSE19027 between smokers and non-smokers samples were used as inputs to the CMap database. Compounds were selected as potential therapeutic agents for smoking-induced LUAD after ranking them according to their negative connectivity enrichment scores.

3. Results

3.1. Identification of DEGs

Based on the cut-off criteria a total of 1092 DEGs were screened, including 313 upregulated and 779 downregulated genes in the smoker samples compared to non-smoker samples.

3.2. Network construction and screening of hub genes

According to acquired information from the STRING database, a total of 966 nodes and 4119 edges were mapped in the network. A significant module (MCODE score >10) including 69 nodes and 364 edges was identified in the PPI network (Figure 1). The top 20 genes were ranked using four calculation algorithms of the CytoHubba plugin including Maximal Clique Centrality (MCC), Maximum Neighborhood Component (MNC), Degree, and Betweenness. Finally, nine intersecting genes (*ACTR2*, *ANAPC10*, *CXCL8*, *CXCR4*, *DAB2*, *GNG11*, *PTGDR2*, *SOCS3*, and *VAMP3*) of the top 20 ranked DEGs were selected as hub genes (Table 1, Figure 2).

Table 1. Top 20 genes evaluated in the PPI network. Intersecting genes are shown in bold.

Gene	MCC	Gene	MNC	Gene	Degree	Gene	Betweenness
GNG11	4037681.0	GNG11	26.0	GNG11	27.0	CXCL8	1692,21
LPAR1	3674790.0	CXCL8	19.0	CXCL8	19.0	GNG11	1153,11
DRD4	3634320.0	COL7A1	18.0	COL7A1	18.0	TGFB1	849,85
CXCL8	3634224.0	LPAR1	17.0	LPAR1	17.0	TGFA	757,71
GNAI3	3633864.0	CXCR4	15.0	CXCR4	15.0	SOCS3	683,25
GRM2	3633842.0	GRM1	14.0	GRM1	14.0	COL7A1	635,85
HTR1D	3633840.0	DRD4	13.0	PTGDR2	13.0	ACTR2	602,41
CXCR4	3629058.0	PTGDR2	13.0	SOCS3	13.0	CXCR4	441,50
PTGDR2	3628814.0	SOCS3	13.0	DRD4	13.0	ANAPC10	363,00
SOCS3	3628808.0	GNAI3	12.0	ADRBK1	12.0	COL2A1	246,51
ANAPC10	3628802.0	GRM2	12.0	GNAI3	12.0	DAB2	212,35
VAMP3	3628802.0	ADRBK1	12.0	GRM2	12.0	ADM	167,87
CLTB	3628802.0	HTR1D	11.0	VAMP3	12.0	SEC24D	155,40
CLTCL1	3628802.0	ANAPC10	11.0	HTR1D	11.0	VAMP3	131,61
ACTR2	3628801.0	CLTB	11.0	ANAPC10	11.0	IL13	91,75
DAB2	3628801.0	CLTCL1	11.0	CLTB	11.0	ADRBK1	71,23
KLHL20	3628800.0	TGFA	11.0	CLTCL1	11.0	PTGDR2	66,04
KCTD7	3628800.0	VAMP3	10.0	ACTR2	11.0	GRM1	65,33
FBXO21	3628800.0	ACTR2	10.0	DAB2	11.0	STX17	60,76
ARIH2	3628800.0	DAB2	10.0	TGFA	11.0	SEC22B	60,76

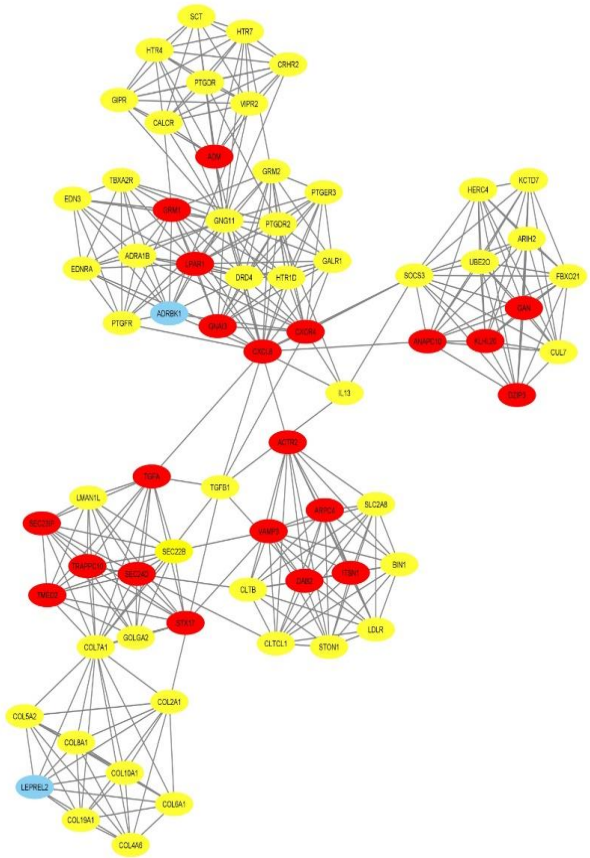

Figure 1. Top module from PPI network. Red nodes represent upregulated genes, and yellow nodes represent downregulated genes.

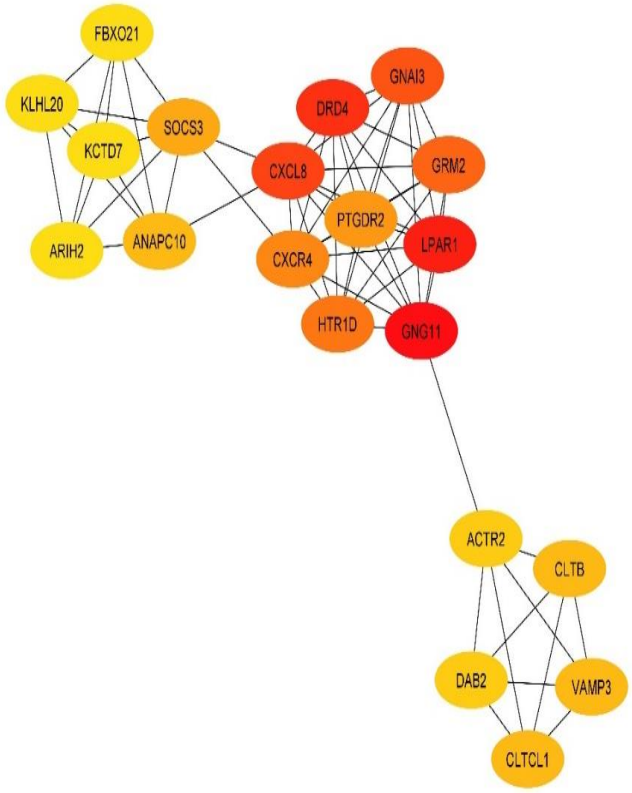

Figure 2. Visualization of the hub genes using cytoHubba plugin. Color grade red to yellow represents MCC scores.

3.3. GO and KEGG pathway enrichment analysis

GO term enrichment results revealed that DEGs are significantly enriched in BP (biological process) and CC (cellular component). As indicated in Table 2, the hub genes were mainly enriched in calcium-mediated signaling, G-protein coupled receptor signaling pathway, chemotaxis, chemokine-mediated signaling pathway, and movement of cell or subcellular component at BP. the hub genes were significantly enriched in clathrin-coated vesicle membrane, intracellular, and clathrin-coated vesicle at CC. In addition, KEGG pathway enrichment results showed that the hub genes were significantly enriched in two pathways including Chemokine signaling pathway and Pathways in cancer (Table 2).

Table 2. KEGG pathway enrichment analysis of hub genes. GO; Gene Ontology, BP; biological process, CC; cellular component, MF; molecular function, KEGG; Kyoto Encyclopedia of Genes and Genomes.

Category	GO Term	p-value
BP	calcium-mediated signaling	2,50E-04
BP	chemotaxis	1,40E-03
BP	G-protein coupled receptor signaling pathway	7,00E-03
BP	chemokine-mediated signaling pathway	3,30E-02
BP	movement of cell or subcellular component	4,00E-02
CC	clathrin-coated vesicle membrane	4,80E-03
CC	intracellular	1,70E-02
CC	clathrin-coated vesicle	2,50E-02
KEGG	Chemokine signaling pathway	1,00E-02
KEGG	Pathways in cancer	4,20E-02

3.4. Validation of hub genes

Prognostic potentials of nine hub genes were evaluated by survival analysis using the Kaplan-Meier plotter and GEPIA. The results demonstrated that among the hub genes only *ACTR2* and *ANAPC10* expression levels significantly associated ($p<0.005$) with the OS of patients with LUAD. Increased *ACTR2* and *ANAPC10* expression could result in a worse OS rate in LUAD patients (Figure 3a-b and d-e). It was also determined that the expressions of these hub genes were higher in LUAD tissues compared to normal tissues (Figure 3c-f). In fact, in the analysis of the GSE19027 dataset, it was also found that the *ACTR2* and *ANAPC10* mRNA levels in smokers increased 3.83-fold and 3.11-fold in smokers compared to non-smokers, respectively. The results indicate that these genes can be effective prognostic factors for LUAD.

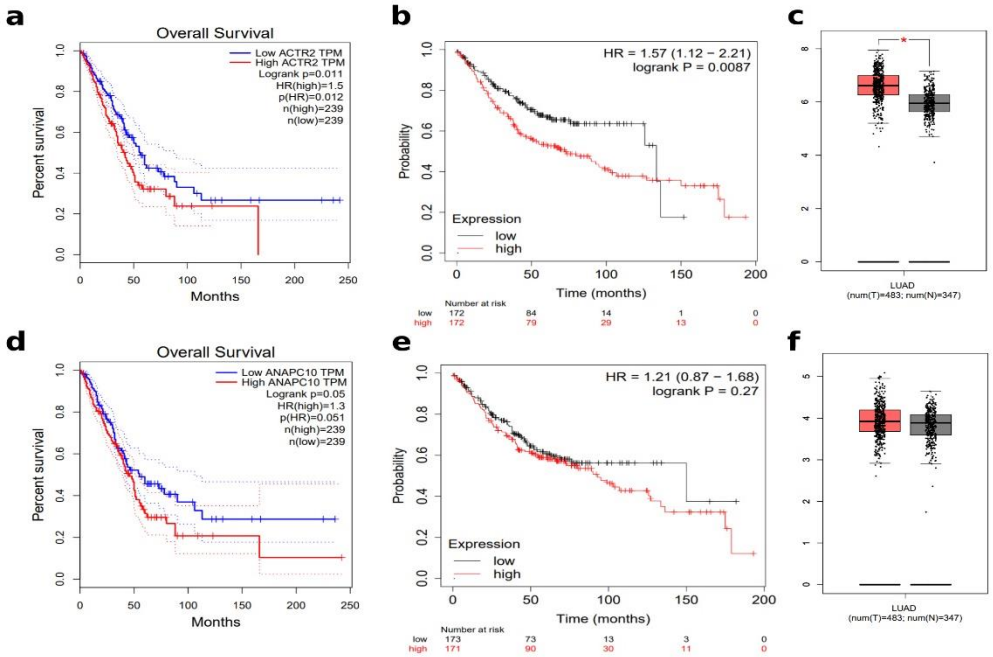

Figure 3. Survival curves and expression boxplots of the hub genes. Overall survival analyses of *ACTR2* using GEPIA **(a)** and Kaplan-Meier plotter **(b)** database. Overall survival analyses of *ANAPC10* using GEPIA **(d)** and Kaplan-Meier plotter **(e)** database. Validation of the expression levels (mRNA) of *ACTR2* **(c)** and *ANAPC10* **(f)** in LUAD samples and normal lung samples using the GEPIA platform.

3.5. Candidate small-molecule drugs screening

To identify candidate small-molecular drugs that could be used to control LUAD progression, prognosis-related DEGs (*ACTR2* and *ANAPC10*) were submitted to the CMap. The related six small molecule drugs with highly significant correlations ($p<0.01$ and higher negative connectivity score) are listed in Table 3. Among these small molecules, MS-275, domperidone, and clomifene showed a higher negative correlation with a smaller p-value.

Table 3. CMap analysis results.

Rank	CMap name	Mean	N	Enrichment	p-value	Specificity	% non-null
1	MS-275	-0,921	2	-0,991	0,00026	0,0617	100
2	domperidone	-0,884	2	-0,981	0,00076	0	100
3	clomifene	-0,878	2	-0,981	0,00082	0,015	100
10	fusaric acid	-0,836	2	-0,956	0,00414	0	100
11	seneciphylline	-0,865	2	-0,951	0,00531	0,0324	100
16	pancuronium bromide	-0,826	2	-0,94	0,00752	0,0216	100

4. Discussion

According to World Health Organization (WHO) reports, non-communicable diseases (NCDs), including cardiovascular diseases, stroke, cancer, and chronic lung disease are responsible for almost 70% of deaths globally [12]. However, it has been shown that almost half of cancer-related deaths can be prevented by modifying lifestyle behaviors or avoiding main risk factors including malnutrition, alcohol consumption, and tobacco use. In addition, early detection of cancer is a highly effective and long-term strategy in reducing the global cancer burden. The most common lethal tumor in the world, lung cancer, causes 1.6 million deaths each year and accounts for 19.4% of the total cancer-related deaths [13]. NSCLC, one of the two main subtypes of lung cancer, comprises 85% of all lung cancers. NSCLC is also classified into three subtypes: large cell carcinoma, squamous cell carcinoma, and adenocarcinoma [14]. Lung adenocarcinoma (LUAD) is the most common type of lung cancer (around 40% of all) [15]. However, underlying mechanisms responsible for the initiation, development, and metastasis of the disease are still poorly understood. Therefore, identifying the molecular drivers associated with LUAD development may contribute to the development of new approaches for early diagnosis and disease management. In this study, a total of 9 significantly dysregulated genes between current smokers without cancer and never smokers were identified in the GEO dataset GSE19027 using bioinformatics analysis. Finally, nine genes were screened out as hub genes. Among them, it was found that overexpression of *ACTR2* and *ANAPC10* were significantly associated with shorter patients' survival.

Different types of mammalian cells, including fibroblasts, hematopoietic cells, and embryonic cells, have their directional motility, also known as cell migration, which plays an essential role in the physiologic functions [16]. However, abnormal cytoplasmic protrusions, such as lamellipodia, can mediate cancerous cells to migrate and metastasize in a coordinated manner in malignant cells [17]. The Arp2/3 (Actin-related protein 2: ACTR2 and 3: ACTR3) complex is responsible for lamellipodium formation and thus involved in the movement of many types of cells. Deletion and RNA-mediated interference (RNAi) studies on Arp2/3 complex have also indicated that this complex is essential for cell migration [18, 19]. Previous studies also demonstrated that the Arp2-positive cells with higher levels of ACTR2 were accumulated within the tumor tissue [20]. Anaphase-promoting complex (APC/C), consisting of 11–13 highly conserved subunits, marks target cell cycle proteins for degradation. ANAPC10, a subunit of the APC/C complex, displays an essential role in substrate recognition [21]. Recent studies, such as that performed by Wang et al. [22] discovered that *ANAPC10* is overexpressed in NSCLC cell lines and promotes the proliferation of cells. They also showed that the knockdown of *ANAPC10* significantly inhibited the migration of NSCLC cells. Taken together, *ACTR2* and *ANAPC10* may be a valuable clinical indicator of lung adenocarcinoma development and progression.

Based on the small-molecule analysis, a set of small-molecule that could reverse smoking-induced abnormal gene expression that can lead to the LUAD development was determined. According to CMap predication, it was found that the drug signatures significantly correlates with *ACTR2* and *ANAPC10*

gene signatures. Among these, MS-275, also known as entinostat is a histone deacetylase inhibitor (HDACi) that increases acetylated histones and leads to transcriptional suppression [23]. Moreover, recent studies also reported that MS-275 potentiated and facilitated inhibitory effects of different antitumor suppressors in lung adenocarcinoma [24]. Furthermore, other noteworthy molecules and bioactive metabolites we found are listed in Table 3.

In summary, this study was designed to identify critical genes that might be involved in the smoking-associated LUAD progression. In addition, a group of small molecules that can increase efficacy in LUAD therapy have been identified. However, future experimental investigations are needed to validate the predicted molecules.

Conflicts of interest

The authors state that there is no conflict of interests.

References

[1] Sung H., Ferlay J., Siegel R.L., Laversanne M., Soerjomataram I., Jemal A., Bray F., Global Cancer Statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries, *CA: Cancer J. Clin.*, 71(3) (2021) 209-249.

[2] Bender E., Epidemiology: the dominant malignancy, *Nature*, 513 (2014) S2-S3.

[3] Hammouz R.Y., Kostanek J.K., Dudzisz A., Witas P., Orzechowska M., Bednarek AK., Differential expression of lung adenocarcinoma transcriptome with signature of tobacco exposure, *J. App. Genet.*, 61 (2020) 421-437.

[4] Herbst R.S., Morgensztern D., Boshoff C., The biology and management of non-small cell lung cancer, *Nature*, 553 (2018) 446-454.

[5] Ducray F., Honnorat J., Lachuer J., DNA microarray technology: principles and applications to the study of neurological disorders, *Rev. Neurol.*, 163 (2007) 409-420.

[6] Sahin U., Derhovanessian E., Miller M., Kloke B.P., Simon P., Löwer M., Bukur V., Tadmor A.D., Luxemburger U., Schrörs B., Personalized RNA mutanome vaccines mobilize poly-specific therapeutic immunity against cancer, *Nature*, 547 (2017) 222-226.

[7] Li L., Lei Q., Zhang S., Kong L., Qin B., Screening and identification of key biomarkers in hepatocellular carcinoma: evidence from bioinformatic analysis, *Oncol. Rep.*, 38 (2017) 2607-2618.

[8] Liu W., Ouyang S., Zhou Z., Wang M., Wang T., Qi Y., Zhao C., Chen K., Dai L., Identification of genes associated with cancer progression and prognosis in lung adenocarcinoma: Analyses based on microarray from Oncomine and The Cancer Genome Atlas databases, *Mol. Genet. Gen. Med.*, 7 (2019) e00528.

[9] Wang X., Pittman G.S., Bandele O.J., Bischof J.J., Liu G., Brothers J.F., Spira A., Bell D.A., Linking polymorphic p53 response elements with gene expression in airway epithelial cells of smokers and cancer risk, *Hum. Genet.*, 133 (2014) 1467-1476.

[10] Tang Z., Li C., Kang B., Gao G., Li C., Zhang Z., GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses, *Nucleic Acids Res.*, 45 (2017) W98-W102.

[11] Győrffy B., Surowiak P., Budczies J., Lánczky A., Online survival analysis software to assess the prognostic value of biomarkers using transcriptomic data in non-small-cell lung cancer, *PLoS One.*, 8 (2013) e82241.

[12] World Health Organization, *Noncommunicable Diseases Country Profiles – WHO Global Report, Geneva: WHO,* (2018).

[13] Forum of International Respiratory Societies. The Global Impact of Respiratory Disease – Second Edition. Sheffi eld, European Respiratory Society (2017).

[14] Zappa C., Mousa S.A., Non-small cell lung cancer: current treatment and future advances, *Transl. Lung Cancer Res.*, 5 (2016) 288-300.

[15] Denisenko T.V., Budkevich I.N., Zhivotovsky B., Cell death-based treatment of lung adenocarcinoma, *Cell Death Dis.*, 9 (2018) 1-14.

[16] Yamaguchi H., Lorenz M., Kempiak S., Sarmiento C., Coniglio S., Symons M., Segall J., Eddy R., Miki H., Takenawa T., Molecular mechanisms of invadopodium formation: the role of the N-WASP–Arp2/3 complex pathway and cofilin, *J. Cell. Biol*., 168 (2005) 441-452.

[17] Kurisu S., Suetsugu S., Yamazaki D., Yamaguchi H., Takenawa T., Rac-WAVE2 signaling is involved in the invasive and metastatic phenotypes of murine melanoma cells, *Oncogene*., 24 (2005) 1309-1319.

[18] Sawa M., Suetsugu S., Sugimoto A., Miki H., Yamamoto M., Takenawa T., Essential role of the C. elegans Arp2/3 complex in cell migration during ventral enclosure, *J. Cell. Sci.*, 116 (2003) 1505-1518.

[19] Hudson A.M., Cooley L., A subset of dynamic actin rearrangements in Drosophila requires the Arp2/3 complex, *J. Cell. Biol*., 156 (2002) 677-687.

[20] Semba S., Iwaya K., Matsubayashi J., Serizawa H., Kataba H., Hirano T., Kato H., Matsuoka T., Mukai K., Coexpression of actin-related protein 2 and Wiskott-Aldrich syndrome family verproline-homologous protein 2 in adenocarcinoma of the lung, *Clin. Cancer Res*., 12 (2006) 2449-2454.

[21] Marrocco K., Criqui M.C., Zervudacki J., Schott G., Eisler H., Parnet A., Dunoyer P., Genschik P., APC/C-mediated degradation of dsRNA-binding protein 4 (DRB4) involved in RNA silencing, *PLoS One*., 7 (2012) e35173.

[22] Wang Y., Han T., Gan M., Guo M., Xie C., Jin J., Zhang S., Wang P., Cao J., Wang J-B. A novel function of anaphase promoting complex subunit 10 in tumor progression in non-small cell lung cancer, *Cell Cycle*., 18 (2019) 1019-1032.

[23] Gerson S.L., Caimi P.F., William B.M., Creger R.J., Pharmacology and molecular mechanisms of antineoplastic agents for hematologic malignancies, Hematology (Seventh Edition), Elsevier, (2018) 849-912.

[24] Luo B.L., Zhou Y., Lv H., Sun S.H., Tang W.X., MS-275 potentiates the effect of YM-155 in lung adenocarcinoma via survivin downregulation induced by miR-138 and miR-195, *Thorac. Can.*, 10 (2019) 1355-1368.