Original Article

Thai Journal of Pharmaceutical Sciences

In silico discovery of potential diagnostic biomarkers of lung cancer

Ruzianisra Mohamed, Raja Nur Hidayah Raja Hasnan

Department of Pharmaceutical Life Sciences, Faculty of Pharmacy, Universiti Teknologi MARA Puncak Alam Campus, 42300 Bandar Puncak Alam, Selangor, Malaysia

Corresponding Author:
Ruzianisra Mohamed, Department of Pharmaceutical Life Sciences, Faculty of Pharmacy, Universiti Teknologi MARA Puncak Alam Campus, 42300 Bandar Puncak Alam, Selangor, Malaysia.
E-mail: ruzianisra@uitm.edu.my

Received: Sep 09, 2019
Accepted: Dec 23, 2019
Published: Jun 05, 2020

ABSTRACT

Introduction: Understanding hub genes implicated in lung cancer (LC) metastasis will help in finding effective ways to diagnose and cure cancer. Accurate identification of protein biomarkers helps in improving the prognosis of LC. Here, the focus of this study was to discern the biomarkers that are implicated in LC. **Materials and Methods:** Three datasets were extracted from gene expression omnibus (GEO) database. GEO2R tool was used to identify the differentially expressed genes (DEGs) between LC and normal lung samples. FunRich software, Enrichr, and Kyoto Encyclopedia of Genes and Genomes database were used to identify the common DEGs in LC, continued by identifying functions and pathways. Next, protein-protein interactions were obtained from search tool for the retrieval of interacting genes database. The hub genes were identified using CytoHubba tool. Then, the prognostic value in the identified genes was verified using LC database in Kaplan-Meier Plotter platform. **Results and Discussion:** A total of 215 downregulated and 84 upregulated were overlapped. A total of ten hub genes such as interleukin-6, matrix metallopeptidase 9, secreted phosphoprotein 1, enhancer of zeste homolog 2, collagen type 1 alpha 1, platelet endothelial cell adhesion molecule 1 (PECAM1), CDK1, VWF, EDN1, and CD34 were selected. Three significant DEGs were identified to be associated with favorable overall survival in LC patients which were PECAM1, EDN1, and von Willebrand factor (VWF). **Conclusion:** Therefore, this study suggests that all the hub genes may be potential biomarkers and treatment target for LC.

Keywords: Bioinformatics, biomarker, hub genes, lung cancer

INTRODUCTION

Lung cancer (LC) is one of the main causes of cancer-related mortality worldwide.[1] It is divided into two types which are primary LC and secondary LC.[2] This disease also classified into small cell LC (SCLC) and non-SCLC (NSCLC).[3] About 80% of LC cases are due to NSCLC includes large cell carcinoma, squamous cell carcinoma, and adenocarcinoma. In Malaysia, LC is the third most common cancer and the mortality remains high because of late diagnosed. However, most of the patient is diagnosed with NSCLC at Stages II–IV; thus, the patient does not have a chance to do the surgery. Patients with chronic LC usually die 1–5 years from diagnosed and the cure rate is <15%.[4] To this date, the prognosis of NSCLC is still poor even though many researches have been done to increase the understanding about LC pathogenesis and the molecular characteristics of NSCLC. The survival rate of the patient has not been improved.[5] Thus, LC novel predictive biomarkers are needed to improve the treatment and increase the accuracy of prognosis.[6]

Over the years, gene expression omnibus (GEO) database archived and distributed high-throughput gene expressions and functions of genomic data sets. Geo has evolved and now extensively used to examine chromatin structure, methylation of genome, genome-protein interaction, and discover potential biomarkers for cancer diagnosis.[7] Furthermore, to overcome inconsistent results from the study, integrated bioinformatics approaches have been widely utilized in cancer research and a few of bioinformatics reports have been published.[8]

In this study, we used three gene expression profiles from the GEO database to identify differentially expressed genes (DEGs) between human LC and normal lung tissue samples. Then, Kyoto Encyclopedia of Genes and Genomes (KEGG) and Gene Ontology (GO) were applied to analyze the functions of all the DEGs. We also used a search tool for the retrieval of interacting genes (STRING) and Cytoscape to construct protein-protein interaction (PPI) network and identify the top hub genes

associated with LC. Finally, Kaplan-Meier (KM) plotter was used to validate the overall survival analyses of the hub genes.

MATERIALS AND METHODS

Identification of Genes Implicated in LC

GEO Database was used to retrieve a few series of datasets about human lung adenocarcinoma.[9] The database showed a total of 8698 series about adenocarcinoma. The series had been reviewed and three datasets of gene expression related to lung adenocarcinoma were downloaded (GSE 31210, GSE 43458, and GSE 10072). The platform of GSE 31210 was based on GPL570 platform ([HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array). Then, GSE 43458 was based on platform GPL6244 ([HuGene-1_0-st] Affymetrix Human Gene 1.0 ST Array [transcript (gene) version]) and lastly, GSE 10072 was based on platform GPL96 ([HG-U133A] Affymetrix Human Genome U133A Array).

Selecting the DEGs in LC Patients

GEO2R tool was used to analyze all the genes in the datasets by calculating the corrected *P* value and |logFC|.[7] DEGs were genes that met the cutoff criteria, corrected $P < 0.05$ and |logFC| ≥1.0. Then, FunRich software was applied to construct Venn diagram to identify the intersection part for all three datasets.[10]

KEGG Pathway and GO Analysis of DEGs

Enrichr online tool was used to learn more about all the function of the DEGs and their pathways.[11] The gene function can be comprised three terms which are biological process (BP), cellular component (CC), and molecular function (MF). Next, KEGG is a collection of databases dealing with biological, genomes, diseases, pathways, drugs, and chemical substances that were used to identify the pathways for DEGs.[12]

PPI Network Screening and Identification of Hub Gene

The STRING database is a biological database which predicts the PPI.[13] In this study, STRING database and Cytoscape software were used to construct PPI network between the DEGs.[14] In Cytoscape, the plugin tool CytoHubba was used to calculate each protein node degree. The hub genes were the gene with 10 or more gene degrees in the PPI network.[15]

Hub Genes Survival Analysis

The KM plotter is an online tool which capable to assess the effect of 54K genes on survival in 21 types of cancer.[16] This includes 6234 breast cancer, 1440 gastric cancer, 2190 ovarian cancer, and 3452 LC. Thus, the KM plotter of LC was used to identify the prognostic value of the hub genes in lung adenocarcinoma patients.

RESULTS

Identification of Genes Implicated in LC

In this study, three gene datasets (GSE31210, GSE43458, and GSE10072) were selected and downloaded from the GEO database. From the datasets, GSE31210 contained 226 LC samples and 20 normal samples, GSE43458 contained 80 LC samples (40 samples of non-smoker and 40 samples of smoker) and 30 normal samples, and GSE10072 contained 58 LC samples (15 samples of non-smoker and 43 samples of smoker) and 49 normal samples [Table 1]. Based on GEO2R analysis, a total of 2331 DEGs were identified from GSE31210, including 1313 downregulated genes and 1018 upregulated genes. The DEGs met the selection criteria which are $P < 0.05$ and |logFC| ≥1.0. Then, in GSE43458 dataset, 810 DEGs were screened, including 598 downregulated genes and 212 upregulated genes. In GSE10072 dataset, 609 DEGs were obtained and among them, 409 were downregulated genes and 200 upregulated genes. Next, we used Venn analysis in FunRich software to find the correlation of DEGs genes. Thus, based on the Venn diagram, 299 DEGs were significantly expressed among all the dataset [Figure 1]. From 299 DEGs, 215 DEGs were significantly downregulated genes, while 84 DEGs were significantly upregulated genes [Table 2].

Functional and KEGG Pathway Analyses of the DEGs

Enrichr was used to perform the GO functional and KEGG pathway analysis for all of the DEGs. The GO analysis results were divided into BP, CC, and the MF. The results show that the downregulated DEGs were mainly enriched in BP, including the regulation of angiogenesis, extracellular matrix assembly, and regulation of vasculogenesis, glomerulus vasculature development, and regulation of nitric oxide biosynthesis process. The upregulated DEGs were enriched in the extracellular matrix organization, regulation of glial cell differentiation, centromeric sister chromatid cohesion, mitotic spindle checkpoint, and spindle assembly checkpoint.

Next, the CC shows that the downregulated DEGs were mainly enriched in an integral component of plasma membrane, G-protein coupled receptor dimeric complex, membrane raft, lamellar body, and platelet alpha granule membrane. The upregulated DEGs in CC show that they mainly expressed in condensed nuclear chromosome kinetochore, spindle, spindle pole, microtubule cytoskeleton, and endoplasmic reticulum lumen.

As for MF, the downregulated DEGs were mainly expressed in amyloid-beta binding, metalloendopeptidase inhibitor

Table 1: The gene expression for gene expression omnibus lung cancer data profile

Reference	Gene expression omnibus	Platform	Normal	Tumor
Kabbout *et al*.[17]	GSE43458	GPL6244	30	80
Landi *et al*.[18]	GSE10072	GPL96	49	58
Yamauchi *et al*.[19]	GSE31210	GPL570	20	226

Table 2: Analysis of differentially expressed genes by gene expression omnibus 2R and FunRich software

Differentially expressed genes	Gene symbol
Downregulated	ABCA3,ABCA8,ACADL,ADAMTS1,ADAMTSL3,ADARB1,ADGRG6,ADGRL2,ADGRL4,ADH1B,ADIRF,AGER,AGTR1, AGTR2,AHNAK,AKT3,ANGPT1,ANOS1,ANXA3,AOC3,AQP4,ARHGAP29,ARHGAP44,ARHGAP6,ARRB1,BCHE, C14orf132,C7,CA2,CA4,CACNA2D2,CALCRL,CAT,CAV1,CD34,CD36,CD93,CDH5,CDO1,CHRDL1,CLDN18,CLIC5, CPA3,CPB2,CRTAC1,CRYAB,CTNNAL1,CXCL2,CYP4B1,DACH1,DCN,DENND3,DES,DNAJB4,DOCK4,DPT,DUOX1, DUSP1,EDN1,EDNRB,EFEMP1,EMCN,EML1,EMP1,EMP2,EPB41L2,ERG,FABP4,FAM105A,FAM107A,FAM189A2, FBLN5,FCN3,FEZ1,FGFR2,FHL1,FHL5,FLI1,FLRT3,FMO2,FOSB,GHR,GIMAP6,GNG11,GPC3,GPM6A,GPM6B,GPRC5A, GRK5,HBB,HBEGF,HEG1,HEY1,HIGD1B,HLF,HOXA5,HSD17B6,HSPB8,ICAM2,ID1,ID4,IL1RL1,IL33,IL6,INPP5A,ITGA8, ITIH5,ITM2A,JAM2,KCNJ15,KIAA1462,KL,KLF4,KLF6,LAMP3,LDB2,LDLR,LHFP,LIMCH1,LMCD1,LMO7,LPL,LRRC32, LRRC36,LRRN3,LYVE1MARCO,MEIS1,METTL7A,MFAP4,MME,MMRN2,MRC1,MSR1,MT1M,MYH10,MYH11,MYL9, NEBL,NEDD4L,NPR1,NPR3,NRN1,OGN,OLFML1,OLR1,P2RY14,P3H2,PALMD,PCDH17,PCOLCE2,PDK4,PDZD2, PECAM1,PGC,PHACTR2,PID1,PIP5K1B,PLA2G1B,PLLP,PTPRB,PTRF,RAB11FIP1,RAMP2,RAMP3,RBP4,RECK,RGCC, S1PR1,SASH1,SCEL,SCGB1A1,SDPR,SEMA5A,SEMA6A,SFTPC,SFTPD,SH3BP5,SLC1A1,SLC39A8,SLC6A4,SLCO2A1, SLIT2,SMAD6,SOCS2,SORBS1,SOSTDC1,SPOCK2,SPTBN1,SRPX,STARD13,STXBP6,SYNE1,TACC1,TBX3,TCF21,TEK, TGFBR2,TGFBR3,THBD,TIE1,TIMP3,TMEM100,TMEM204,TMEM47,TNNC1,TNS1,TPPP3,TSPAN7,VIPR1,VSIG4, VWF,WASF3,WIF1ZBTB16
Upregulated	KIF4A,KIF11,MUC16,MXRA5,MUC5B,MMP1,MMP11,LCN2,MELK,LGSN,LGR4,MMP12,MMP7,MMP9,TCN1,SULF1, THBS2,TFAP2A,SPINK1,SPP1,STEAP1,TIMP1,TYMS,TTK,TNFRSF21,TMPRSS4,TMPRSS11E,TPX2,TOX3,TOP2A,PLAU, NQO1,PCP4,PBK,SLC2A1,SLC7A11,RRM2,SFN,S100P,CEP55,CENPF,CFB,CHI3L1,CDH3,CDK1,CEACAM5,CDKN3, CP,COMP,CRABP2,CST1,COL3A1,COL1A1,COL11A1,COL10A1,ASPM,ADAM28,ABCC3,AGR2,CCNB2,CD24,CDC20, CCNB1,BUB1B,BUB1,CXCL13,GREM1,GPR87,KIAA0101,KDELR3,KCNN4,HMGB3,IGF2BP3,EZH2,FAP,DLGAP5, CYP24A1,ECT2,DSP,GINS1,GDF15,GCNT3,GOLM1,GALNT7

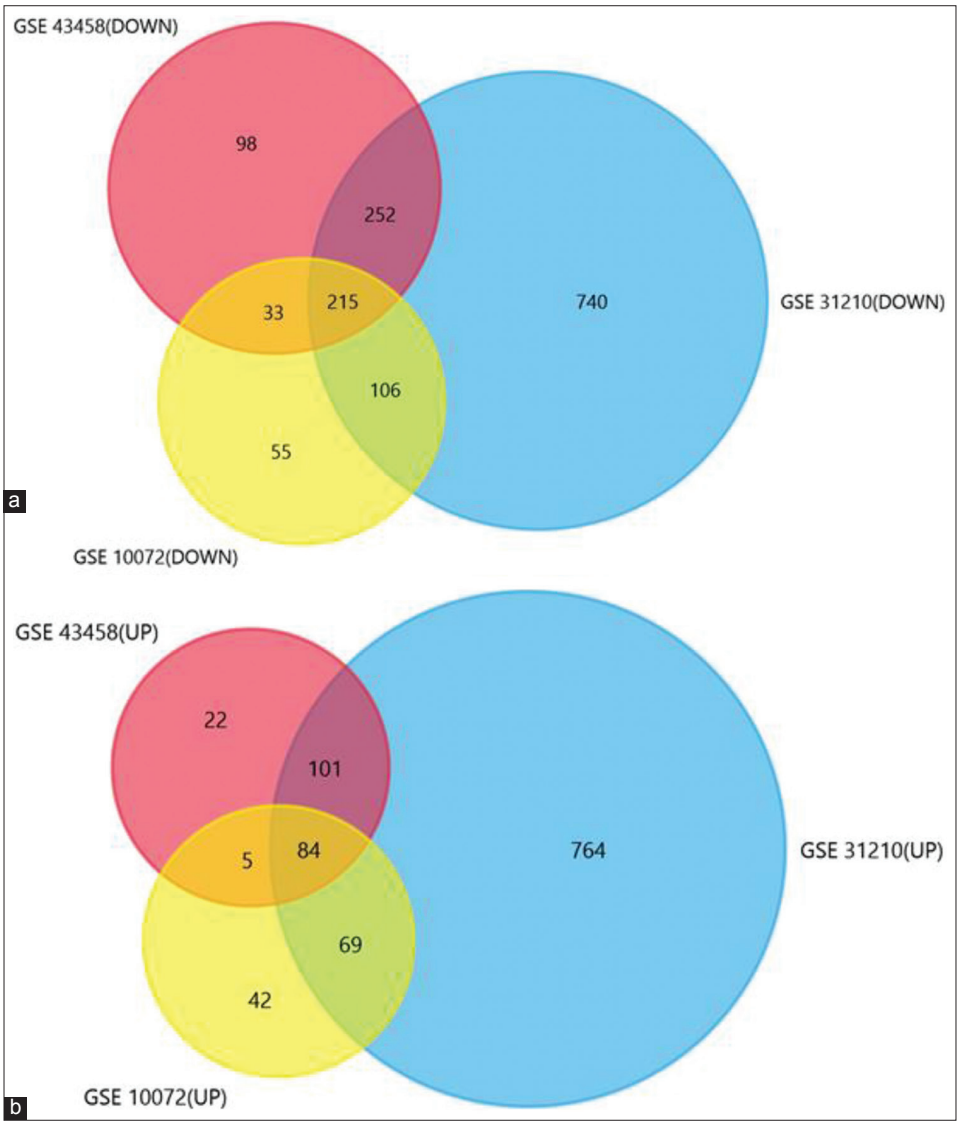

Figure 1: The Venn analysis which shows the correlation between all three gene expression omnibus datasets. Venn diagram (a) shows the correlation of downregulated differentially expressed genes (DEGs) while Venn diagram (b) shows the correlation between upregulated DEGs

activity, transmembrane receptor protein serine, transforming growth factor-beta activated receptor activity, and low-density lipoprotein particle binding. In addition, the analysis also showed that the upregulated DEGs were mainly enriched in serine-type peptidase activity, peptidase activity, metalloendopeptidase activity, histone kinase activity, and metallopeptidase activity.

The KEGG pathway for downregulated DEGs was mainly expressed for vascular smooth muscle contraction, AGE-RAGE signaling pathway in diabetic complication, PPAR signaling pathway, tight junction, and cell adhesion molecules. The upregulated DEGs significantly enriched in cell cycle, p53 signaling pathway, oocyte meiosis, ECM-receptor interaction, and focal adhesion.

PPI Network Screening and Identification of Hub Gene

STRING database was used to construct the PPI network among the DEGs. Then, Cytoscape software was used to identify the hub gene. The protein interaction networks had a total of 298 nodes and 1177 edges that were involved in the network, as showed in Figure 2. Based on the connectivity degree in the PPI network, the top 10 genes were evaluated [Figure 3]. Thus, based on these results, interleukin 6 (IL6) had the highest-ranked in degree of connectivity which was 79, followed by matrix metallopeptidase 9 (MMP9, degree =51), secreted phosphoprotein 1 (SPP1, degree =37), histone-lysine N-methyltransferase enhancer of zeste homologue 2 (EZH2, degree =34), alpha-1 type 1 collagen (COL1A1, degree =33), platelet and endothelial cell adhesion molecule 1 (PECAM1, degree =32), cyclin-dependent kinase 1 (CDK1,degree =31), von Willebrand factor (VWF, degree =31), endothelin 1 (EDN1, degree =30), and transmembrane phosphoglycoprotein (CD34, degree =30). All these genes may play important roles in the development of lung adenocarcinoma.

The Hub Genes Survival Analysis

KM plot was constructed for each of the ten hub genes to identify their prognostic value. On the platform, there were 1926 LC patients involved in the overall survival analysis. From the KM plot, seven significant DEGs which were IL6, MMP9, SPP1, EZH2, COL1A1, CDK1, and CD34 were identified to be related to discouraging overall survival in LC patients. Furthermore,

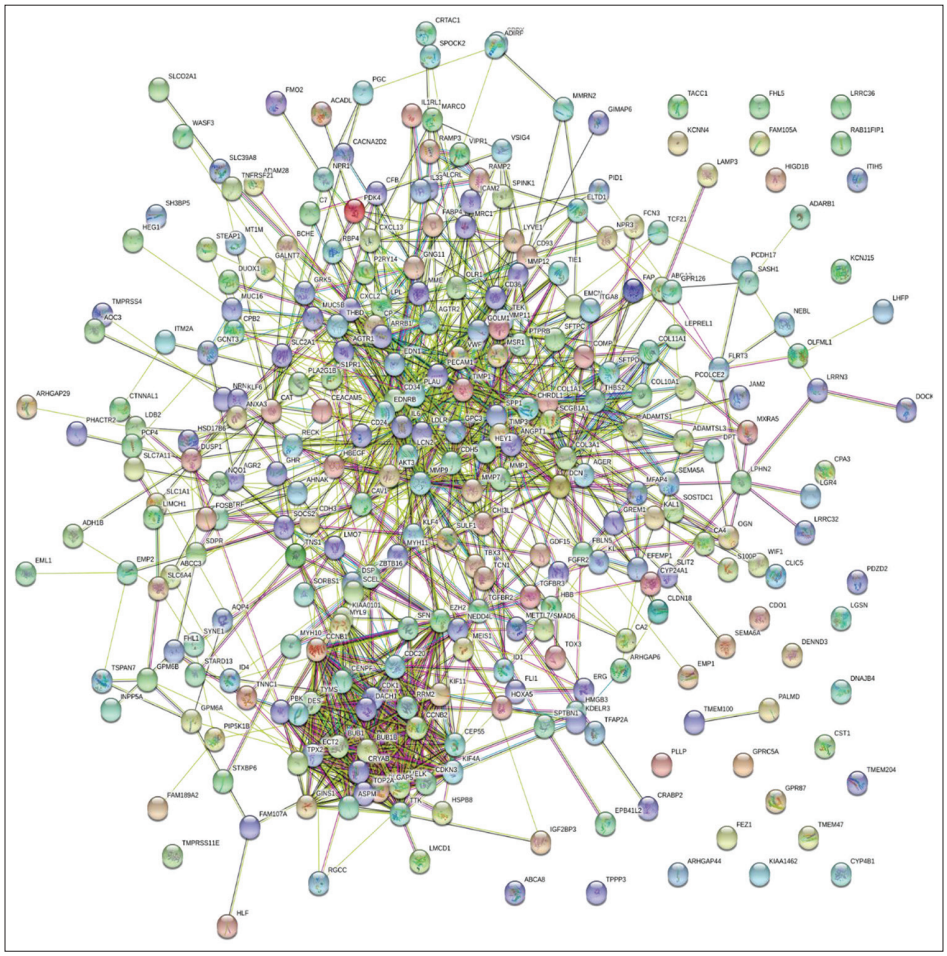

Figure 2: The search tool for the retrieval of interacting genes protein-protein interaction networks of 215 downregulated and 84 upregulated genes. The network had 298 nodes and 1177 edges. The lines represent the interaction of protein between the genes, the circles represent the genes, and the results within the circles show the protein structure. The network has significantly more interaction because the proteins have more interactions among themselves that what would be expected for a random set of proteins of similar size, drawn from genome. The enrichment indicates that the proteins are at least partially biologically connected, as a group

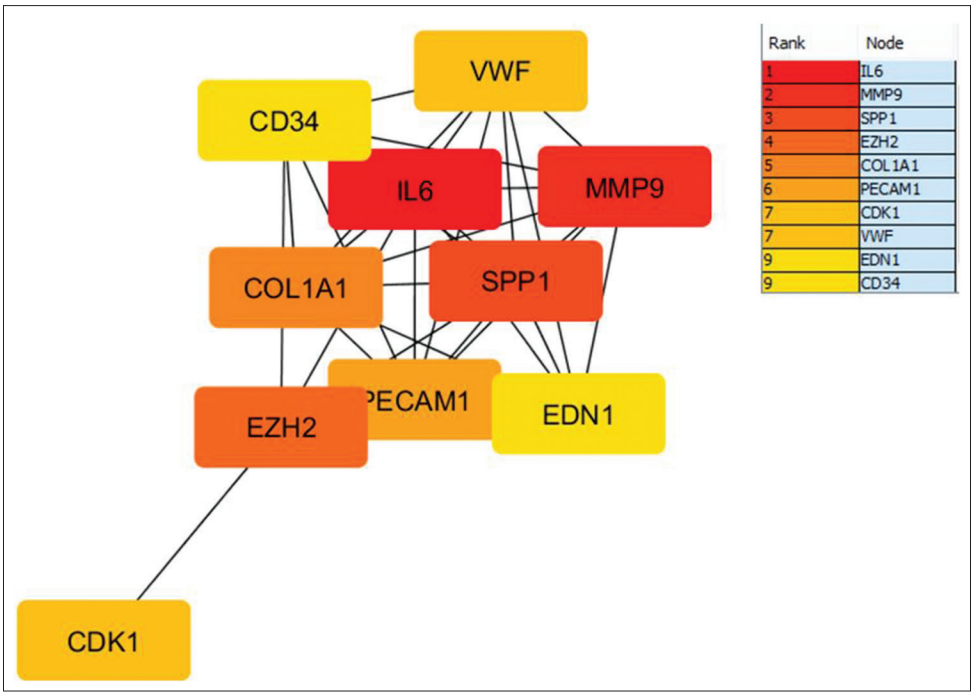

Figure 3: Top ten hub genes subnetwork from protein-protein network using Cytoscape software. The different color of nodes represents the degree of connectivity. The color shows the hub genes rank from 1 to 10. Yellow color represents the lowest degree, orange color shows the intermediate degree, and red color shows the highest degree among all the hub genes

three significant DEGs were identified to be associated with favorable overall survival in LC patients which were PECAM1, EDN1, and VWF.

DISCUSSION

In this study, 215 downregulated and 84 upregulated DEGs were identified between the normal tissue and the LC tissue. The results were obtained after screening three datasets from GEO database. Then, all DEGs were divided according to their GO functional activity such as CC, BP, and MF and KEGG pathway. The results showed that the LC patients' DEGs involved in extracellular matrix organization, G-protein coupled receptor dimeric complex, and amyloid-beta binding. The analysis of KEGG pathway also showed that the DEGs were enriched in vascular smooth muscle contraction, PPAR signaling pathway, tight junction and cell adhesion molecules, p53 signaling pathway, oocyte meiosis, ECM-receptor interaction, and focal adhesion. Next, STRING database and Cytoscape were used to construct the PPI network and to screen 10 hub genes which are IL-6, MMP9, SPP1, EZH2, COL1A1, PECAM1, CDK1, VWF, EDN1, and CD34. Based on the results, five genes are upregulated (COL1A1, CDK1, EZH2, MMP9, and SPP1) and another five genes (CD34, EDN1, IL-6, PECAM1, and VWF) are downregulated in LC patients. Moreover, KM plotter was also used to identify the effect of the hub genes on the overall survival in LC patients.

IL-6 is a gene that encodes a cytokine which functions in the maturation and inflammation of β cells. This gene function is associated with various inflammation disease condition and it differentiates β cells into Ig secreting cells which participate in the differentiation of monocyte and lymphocyte. IL-6 is secreted by a wide variety of cells include the tumor cell. It is found to be elevated in tumor and serum tissue of cancers such as LC, breast cancer, colorectal cancer, prostate cancer, and cervical cancer.[20] According to a study by Silva and colleague,[21] IL-6 was identified to be related to gender, showing a higher level in male. Shang *et al*. reported that[22] IL-6 regulates the proliferation of LC. It is positively correlated with the distant metastasis and lymph node metastasis in NSCLC. Therefore, the result shows that IL-6 also could stimulate metastasis in LC. Based on the result in this study, patients with high-level expression of IL-6 have unfavorable value for overall survival in NSCLC.

In normal physiological processes, MMP9 involves in the breakdown of the extracellular matrix, including reproduction, embryonic development, and tissue remodeling. Most of MMP's are inactive proproteins which will be activated when it is cleaved by extracellular proteinases. The gene that is encoded by this enzyme degrades Type IV and V collagens. MMP9 has been identified to relate to the pathology of cancers such as migration, angiogenesis, and metastasis. MMP9 can cause cancer to develop and progress and it is a significant target for several types of cancers such as NSCLC, giant cell tumor of bone, cervical cancer, ovarian cancer, and breast cancer.[23] MMP9 gene expression is present in the normal lung tissue, but the LC tissue had a significantly higher expression. This study result was in agreement with El-Badrawy *et al*.[24] who reported that MMP9 may act as a potential biomarker in LC. It could also help in differentiating the types of LC. This is due to MMP9 has a higher level of expression in NSCLC compared to other types of LC.

The SPP1 encoded a protein and participated in the attachment of osteoclast with a mineralized bone matrix. The

encoded protein binds tightly to hydroxyapatite. SPP1 gene also plays a role as a cytokine which involved in activating the production of IL-12 and interferon-gamma. It also reduces the production of IL-10. SPP1 also promotes tumor metastasis and growth. Recently, it has been reported that SPP1 promotes survival of cancer cell and control the tumor cell related to angiogenesis and inflammation pathway.[25] Moreover, SPP1 has been previously shown that it activates the nuclear factor-kB (NF-kB) in NSCLC.[26]

EZH2 is an important member of polycomb repressive complex which is responsible for the repression of selected gene expression and histone methylation.[27] According to a study by Lopci and Rossi,[28] EZH2 also plays a key role in cancer progression and tumorigenesis. It is also the regulator of tumor angiogenesis. EZH2 was found overexpressed in several cancer, including gastric cancer, breast cancer, prostate cancer, colorectal cancer, and LC. A lot of studies have evaluated the EZH2 overexpression may be a prognostic factor for survival in LC patients.[29] In this research, EZH2 expression has been shown to have unfavorable value in the overall survival of LC. However, more studies need to be conducted to get a more conclusive result.

Chain COL1A1 is a member of Group 1 collagen. It encodes the pro-alpha1 chains of type 1 collagen. In NSCLC tumor, the expression level of COL1A1 messenger RNA was higher than in normal tissue. However, it is not associated with the metastasis stage of tumor node.[30] COL1A1 promoter showed hypermethylation in NSCLC.[31]

The PECAM1 belongs to the immunoglobulin superfamily which important in thrombosis and angiogenesis.[32] PECAM1 also has been reported involved in cell-cell adhesion through heterophilic and homophilic interaction and transduces the intracellular signal. PECAM1 also plays an important role in leukocytes recruitment at inflammatory sites, development in cardiovascular, and the release of bone marrow leukocytes.[33] According to Kuang *et al*.,[32] PECAM1 could be a potential biomarker in LC. Notably, in this study, the KM plotter showed that overexpression of PECAM1 is a favorable prognostic factor for the overall survivor in LC patients.

CDK1 gene encodes a member of the Ser/Thr protein kinase family. The protein known as M-phase promoting factor is important for the transition of eukaryotic cell cycle. It acts as regulatory subunits. Moreover, CDK1 is important for the proliferation and progression of cell cycle and it could dysregulate the CDK1 activity. A study showed that CDK1 was an unfavorable prognostic biomarker for NSCLC.[34] The high level of expression indicates the patients with a high risk of recurrence cancer and poor survival. However, they are also a study associated with CDK1 which identifies that CDK1 inhibition and deprivation of iron are one of the potential strategies to suppress LC.[35]

VWF is a multimeric glycoprotein that is important in primary hemostasis and allowing platelet adhesion to expose subendothelium. This role suggests the potential of cancer metastasis development. The tumor cells will interact with the vessel wall and platelet to extravasate from circulation. A study by Bauer *et al*.[36] reported that VWF in tumor vessels promote tumor which related to thromboembolism and metastasis. VWF will cause platelet aggregation and provoke the coagulation in cancer patients. In addition, VWF preferentially overexpressed in the tumor vasculature of LC compared to other neighboring tissue vasculature. According to a report about VWF,[37] high level of VWF expression was found in the LC patient tissue which in agreement with the results in this study. Moreover, VWF overexpression is favorable in the overall survival of LC patients.

EDN1, a member of the endothelin family, has highly potent vasoconstrictive peptides. Abnormal expression of this gene may cause tumorigenesis. A genotypic polymorphism study showed that EDN1 has a significant relationship with EDNRA genetic polymorphism and presents the severity of LC. The EDN1 gene also causes the development and progression of LC.[38] Moreover, EDN1 has been identified to initiate tumor growth by promoting angiogenesis. The overexpression of EDN1 was observed in several of tumor.[39] In this study, the results of KM plotter showed that overexpression of EDN1 is a favorable factor of overall survival in LC patients.

As discussed above, many previous studies have shown that all the hub genes are linked to LC. Thus, additional bioinformatics analysis and molecular biological experiments are needed to verify our findings.

CONCLUSION

We identified some important genes and systematically showed the pathways and BPs that closely related to LC development. The results gave more comprehensive clues to understand the pathogenesis of LC in patients and identify potential diagnostic biomarker. Notably, compared to other studies, analyzing the data on the hub genes yielded an almost consistent conclusion. The difference on the results was partly due to unequal sample sizes. Further study is needed to verify the findings and assess the hub genes effect.

ACKNOWLEDGMENTS

Special thanks to Universiti Teknologi MARA (UiTM) for the facilities in the Bioinformatics Unit.

CONFLICTS OF INTEREST

The authors declare that there are no conflicts of interest.

REFERENCES

1. Ghadimi K, Bahrami N, Fathi M, Farzanegan B, Naji T, Emami M, *et al*. Diagnostic value of LunX mRNA and CEA mRNA expression in pleural fluid of patients with non-small cell lung cancer. Minerva Pneumol 2017;56:90-5.
2. Lemjabbar-Alaouia H, Hassana O, Yanga YW, Buchanana P. Lung cancer: Biology and treatment options. Biochim Biophys Acta 2015;1856:189-210.
3. Semenova EA, Nagel R, Berns A. Origins, genetic landscape, and emerging therapies of small cell lung cancer. Genes Dev 2015;29:1447-62.
4. Zhou X, Wen W, Shan X, Zhu W, Xu J, Guo R, *et al*. A six-microRNA panel in plasma was identified as a potential biomarker for lung adenocarcinoma diagnosis. Oncotarget 2016;8:6513-25.
5. Grimolizzi F, Monaco F, Leoni F, Bracci M, Staffolani S, Bersaglieri C, *et al*. Exosomal miR-126 as a circulating biomarker

in non-small-cell lung cancer regulating cancer progression. Sci Rep 2017;7:1-12.
6. Kiselev Y, Andersen S, Johannessen C, Fjukstad B, Olsen KS, Stenvold H, *et al*. Transcription factor PAX6 as a novel prognostic factor and putative tumour suppressor in non-small cell lung cancer. Sci Rep 2018;8:1-11.
7. Clough E, Barrett T. The gene expression omnibus database. Methods Mol Biol 2016;1418:93-110.
8. Ni M, Liu X, Wu J, Zhang D, Tian J, Wang T. Identification of candidate biomarkers correlated with the pathogenesis and prognosis of non-small cell lung cancer via integrated bioinformatics analysis. Front Genet 2018;9:469.
9. Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, *et al*. NCBI GEO: Archive for functional genomics data sets--update. Nucleic Acids Res 2013;41:D991-5.
10. Pathan M, Keerthikumar S, Ang CS, Gangoda L, Quek CY, Williamson NA, *et al*. FunRich: An open access standalone functional enrichment and interaction network analysis tool. Proteomics 2015;15:2597-601.
11. Kuleshov MV, Jones MR, Rouillard AD, Fernandez NF, Duan Q, Wang Z, *et al*. Enrichr: A comprehensive gene set enrichment analysis web server 2016 update. Nucleic Acids Res 2016;44:W90-7.
12. Kanehisa M, Goto S, Sato Y, Furumichi M, Tanabe M. KEGG for integration and interpretation of large-scale molecular data sets. Nucleic Acids Res 2012;40:D109-14.
13. Szklarczyk D, Morris JH, Cook H, Kuhn M, Wyder S, Simonovic M, *et al*. The STRING database in 2017: Quality-controlled protein-protein association networks, made broadly accessible. Nucleic Acids Res 2017;45:D362-8.
14. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, *et al*. Cytoscape : A software environment for integrated models of biomolecular interaction networks. Genome Res 2003;13:2498-504.
15. Chin CH, Chen SH, Wu HH, Ho CW, Ko MT, Lin CY. cytoHubba: Identifying hub objects and sub-networks from complex interactome. BMC Syst Biol 2014;8:S11.
16. Kishore J, Goel M, Khanna P. Understanding survival analysis: Kaplan-Meier estimate. Int J Ayurveda Res 2010;1:274-8.
17. Kabbout M, Garcia MM, Fujimoto J, Liu DD, Woods D, Chow CW, *et al*. ETS2 mediated tumor suppressive function and MET oncogene inhibition in human non-small cell lung cancer. Clin Cancer Res 2013;19:3383-95.
18. Landi MT, Dracheva T, Rotunno M, Figueroa JD, Liu H, Dasgupta A, *et al*. Gene expression signature of cigarette smoking and its role in lung adenocarcinoma development and survival. PLoS One 2008;3:e1651.
19. Yamauchi M, Yamaguchi R, Nakata A, Kohno T, Nagasaki M, Shimamura T, *et al*. Epidermal growth factor receptor tyrosine kinase defines critical prognostic genes of stage i lung adenocarcinoma. PLoS One 2012;7:e43923.
20. Kumari N, Dwarakanath BS, Das A, Bhatt AN. Role of interleukin-6 in cancer progression and therapeutic resistance. Tumor Biol 2016;37:11553-72.
21. Silva EM, Mariano VS, Pastrez PR, Pinto MC, Castro AG, Syrjanen KJ, *et al*. High systemic IL-6 is associated with worse prognosis in patients with non-small cell lung cancer. PLoS One 2017;12:e0181125.
22. Shang GS, Liu L, Qin YW. IL-6 and TNF-α promote metastasis of lung cancer by inducing epithelial-mesenchymal transition. Oncol Lett 2017;13:4657-60.
23. Huang H. Matrix metalloproteinase-9 (MMP-9) as a cancer biomarker and MMP-9 biosensors: Recent advances. Sensors (Switzerland) 2018;18:5-7.
24. El-Badrawy MK, Yousef AM, Shaalan D, Elsamanoudy AZ. Matrix metalloproteinase-9 expression in lung cancer patients and its relation to serum MMP-9 activity, pathologic type, and prognosis. J Bronchol Interv Pulmonol 2014;21:327-34.
25. Zhang W, Fan J, Chen Q, Lei C, Qiao B, Liu Q. SPP1 and AGER as potential prognostic biomarkers for lung adenocarcinoma. Oncol Lett 2018;15:7028-36.
26. Psallidas I, Stathopoulos GT, Maniatis NA, Magkouta S, Moschos C, Karabela SP, *et al*. Secreted phosphoprotein-1 directly provokes vascular leakage to foster malignant pleural effusion. Oncogene 2013;32:528-35.
27. Beltran H. EZH2 drives and defines an epigenetic subclass of lung cancer. Sci Transl Med 2016;8:347ec110.
28. Lopci E, Rossi S. Tumor metabolism and prognostic role of EZH2 in non-small cell lung cancer. Transl Cancer Res 2017;6:S982-8.
29. Wang X, Zhao H, Lv L, Bao L, Wang X, Han S. Prognostic significance of EZH2 expression in non-small cell lung cancer: A meta-analysis. Sci Rep 2016;6:19239.
30. Fang S, Dai Y, Mei Y, Yang M, Hu L, Yang H, *et al*. Clinical significance and biological role of cancer-derived Type I collagen in lung and esophageal cancers. Thorac Cancer 2019;10:277-88.
31. Oleksiewicz U, Liloglou T, Tasopoulou KM, Daskoulidou N, Gosney JR, Field JK, *et al*. COL1A1, PRPF40A, and UCP2 correlate with hypoxia markers in non-small cell lung cancer. J Cancer Res Clin Oncol 2017;143:1133-41.
32. Kuang BH, Wen XZ, Ding Y, Peng RQ, Cai PQ, Zhang MQ, *et al*. The prognostic value of platelet endothelial cell adhesion molecule-1 in non-small-cell lung cancer patients. Med Oncol 2013;30:536.
33. Abraham V, Cao G, Parambath A, Lawal F, Handumrongkul C, Debs R, *et al*. Involvement of TIMP-1 in PECAM-1-mediated tumor dissemination. Int J Oncol 2018;53:488-502.
34. Shi YX, Zhu T, Zou T, Zhuo W, Chen YX, Huang MS, *et al*. Prognostic and predictive values of CDK1 and MAD2L1 in lung adenocarcinoma. Oncotarget 2016;7:85235-43.
35. Kuang Y, Guo W, Ling J, Xu D, Liao Y, Zhao H, *et al*. Iron-dependent CDK1 activity promotes lung carcinogenesis via activation of the GP130/STAT3 signaling pathway. Cell Death Dis 2019;10:297.
36. Bauer AT, Suckau J, Frank K, Desch A, Goertz L, Wagner AH, *et al*. Von Willebrand factor fibers promote cancer-associated platelet aggregation in malignant melanoma of mice and humans. Blood 2015;125:3153-63.
37. Xu Y, Pan S, Liu J, Dong F, Cheng Z, Zhang J, *et al*. GATA3-induced vWF upregulation in the lung adenocarcinoma vasculature. Oncotarget 2017;8:110517-29.
38. Lee YL, Shih CM. Association of the endothelin-1 and endothelin a receptor genetic polymorphisms and non-small cell lung cancer in Taiwan. In: 111 Lung Cancer. Lausanne, Switzerland: European Respiratory Society; 2016. p. PA2856.
39. Lu JW, Liao CY, Yang WY, Lin YM, Jin SL, Wang HD, *et al*. Overexpression of endothelin 1 triggers hepatocarcinogenesis in zebrafish and promotes cell proliferation and migration through the AKT pathway. PLoS One 2014;9:e85318.