



Identification of Biomarkers in Key Gene Prediction in Lung Carcinoma

Venkataramanan Swaminathan¹(✉), Tamilambikai Parandaman¹, Kavitha Kannan²,
Norfatiha Binti Bawahi¹, and K. M. Kumar³

¹ Department Diagnostic and Allied Health Science, Faculty of Health Sciences,
Management and Science University (MSU), Shah Alam, Malaysia

s_venkataramanan@msu.edu.my

² School of Bioscience and Technology (SBST), Vellore Institute of Technology (VIT), Vellore,
India

³ School of Life Sciences, Department of Bioinformatics, Pondicherry University, Kalapet,
Pondicherry, India

Abstract. Lung adenocarcinoma is an imminent principal cancer that causes a huge number of mortality for both men and women because of the respiratory epithelium. The survival rate analysis of patients improved year by year by using bioinformatics tools. The purpose of this research aimed to investigate the PPI network of lung cancer and identify the three gene ontology (GO) of gene expression. The discovery of hub gene biomarkers, on the other hand, aids in the investigation of overall survival and the occurrence of lung cancer expression. The GSE176348 and GSE85841 were obtained from GEO Databank, and by analyzing GEO2R, both data classify to the upregulated and downregulated. The regulated data was analyzed by DAVID server. By using server STRING, Cytoscape and CytoHubba, PPI network and hub genes of the upregulated and downregulated were constructed. The OS and expression level were identified by entering both genes to the KM plotter server and GEPIA2. 2 DEGs dataset was get from analyzed logFC value using GEO2R. Results obtained 2 DEGs dataset by GEO2R based on logFC value and the DEGs will present pathway enrichment analysis. The PPI network and hub genes identification results shows by cut-off > 0.9 value of ADH1B, CAV1, GSTA1, ADH1C ADH1A, CXCL12, FGF1, PPARG, FGF2, and IL1A. The analyzed result from OS showed hazard ratio but expression level presented the different gene of both LUSC or LUAD and normal tissues. Useful for understanding biomarker hub genes of the disease and providing bioinformatics tools for prognosis prediction analysis.

Keywords: Lung Cancer · PPI network · DEGs · Hub genes · STRING · Cytoscape · GEPIA2

1 Introduction

Since 1998 in both men and women, lung cancer is the leading cause of cancer death worldwide, and its prevalence has increased in recent decades [1]. In 2020, an estimated 19.3 million new cancer cases will indeed be diagnosed worldwide, with over 10.0

© The Author(s) 2023

R. Somashekhar et al. (Eds.): ICBDS 2022, AHSR 58, pp. 270–293, 2023.

https://doi.org/10.2991/978-94-6463-164-7_19

million cancer deaths [2]. Breast cancer in women has overtaken lung cancer as the most often diagnosed cancer in women [3]. Lung cancer, with an estimated 1.8 million fatalities, remained the largest cause of cancer death (18 percent) [4]. The cells of the respiratory epithelium are the cause of lung cancer [5]. There are two types of small cell lung cancers (SCLC) and non-small cell lung cancers (NSCLC). SCLC accounts for roughly 10% to 15% of all lung malignancies. This form of lung cancer is the most aggressive and fastest-growing of all [6]. Cigarette smoking is highly linked to SCLC. SCLCs spread quickly throughout the body, and they are usually detected after they have spread widely. The most common type of lung cancer is non-small cell lung cancer (NSCLC), which accounts for roughly 85% of all occurrences [7]. NSCLC is divided into three categories based on the cells detected in the tumour. Furthermore, NSCLC patients are divided into three pathologic groups based on their histological subtypes: lung adenocarcinoma (LUAD), large cell carcinoma, and lung squamous cell carcinoma (LUSC) [8]. When it comes to lung squamous cell carcinoma (LUSC) and small cell lung cancer (SCLC), it has a more positive connotation than lung adenocarcinoma (LUAD). Adenocarcinoma is most commonly detected in the outer parts of the lung layers, and it is more likely to be discovered before it has spread. People with adenocarcinoma in situ (formerly known as bronchioloalveolar carcinoma), a kind of adenocarcinoma, have a better prognosis than those with other types of lung cancer.

There are defined grading standards for most epithelial neoplasms that have prognostic relevance. Grading of lung adenocarcinoma is described by the World Health Organization (WHO) as a qualitative assessment of tumour differentiation based on both the extent to which the tumour's architectural pattern reflects normal lung tissue and cytologic atypia [9]. When compared to other types of cancer patients, lung carcinoma patients have significant symptomatology. Because lung cancer patients have a low survival rate in the first year after diagnosis, concerns like palliation, death, and dying are frequently discussed at the time of diagnosis [10]. Patients with Stage I (T1 or T2, N0, M0) illness have a 5-year survival rate of less than 70%. It's possible that earlier identification of peripheral lung cancer will lead to a lower mortality rate [11].

The purpose of this research aimed to investigate the PPI network of lung cancer and to identify the biological component that contains in the gene of the cancer disease. The discovery of hub gene biomarkers, on the other hand, aids in the investigation of overall survival and the occurrence of cancer expression. As a result, bioinformatics tools show a function that contributes to the lung cancer disease analysis of the prognosis, early diagnosis and survival rates.

2 Methodology

Retrieval of data from NCBI website

To carry out this study, extracted the data from the NCBI website (<https://www.ncbi.nlm.nih.gov/gds>). The data extracted was about lung carcinoma under GEO Datasets from the study type of expression profiling by array and specified as homo sapiens. From 808

search results, chosen two different studies which are GEO accession GSE176348 and GSE85841 where each studies contains 24 samples and 16 samples respectively.

Data Processing and DEGs Selection

The GEO2R tool, which is integrated with GEOquery, Bioconductor/R, and the Limma package, was performed and result of analysis GEO data expression that already performed was written by using a table ordering gene using GEO Profiles graphic visualization [12]. Both the chosen studies were run in GEO2R analysis which was available in NCBI. The DEGs screened correspondingly principal standard $p < 0.05$ and $\log_{2}FC > 1$ from the datasets. The value of $\log_{2}FC$ was used to classify the data to upregulated and downregulated DEGs which are negative value is downregulated, while positive value is upregulated.

Identification of common genes via Venn diagram tool

To find out the common genes among the datasets retrieved using Venn diagram website (<http://bioinformatics.psb.ugent.be/webtools/Venn/>) [14]. Venn allows input of gene lists (up to 2 lists, about 88 genes in upregulated and 183 genes in downregulated respectively) and create a Venn diagram to compare the genes. Here, the upregulated genes were input in List 1. The gene sets of downregulated was input into List 2. The overlapping sections of the Venn diagram represent the common genes among the lists and was recorded for analysis.

Functional Enrichment Analysis

DAVID servers (<https://david.ncifcrf.gov/tools.jsp>) affiliated Gene Ontology function and KEGG pathway enrichment with functional annotation. GO examines the function in large amount of genomic or transcription data, where as KEGG pathways examine data related to the integration of molecules or genes [15]. DAVID v6.8 analysed and spot a biological datasets and analysis tool to extracting biological events from the gene lists. DEGs of the enrichment pathway, molecular, and biological function GO with adjusted significant value (p -values 0.05 and FDR0.05) were investigated.

PPI Network Analysis using STRING

The PPI network is used by the STRING online database (<https://string-db.org/>) to compare DEGs based on the dataset's results. To avoid erroneous PPI network as strong achieved for PPI network, cut off > 0.9 was fixated as a high confidence interaction score. The PPI network was consistently removed by the cut-off.

Construction of PPI Network and Hub Genes Identification Using Cytoscape (V 3.9.0)

Cytoscape software v 3.9.0, which was obtained from the STRING database, was used to visualize the PPI network. The high-level data was provided by the STRING database with Cytoscape software to project data, which demonstrated the network of enrichment network and functional detection [16]. The hub gene was generated using cytoHubba in Cytoscape (v.3.9.0), which found the top five hub genes that were for upregulated and

downregulated of DEGs respectively and extracted the sub network, where as the MCC had the highest performance in constructing the PPI network.

The Validation of Survival Analysis Hub Genes

The Kaplan-Meier plotter (<http://kmplot.com/analysis/index.php?p=service&cancer=lung>) was done to analyse the effect of 54 000 genes the mRNA, miRNA, and protein on survival based on 21 types of cancer counting breast cancer, ovarian cancer, lung cancer, and gastric cancer [17]. As shown in the survival figure, the hazard ratio (HR) with a 95 percent confidence interval and log-rank P were determined using the R and Bioconductor package. The GEO database was used to identify overall survival (OS) and relapse-free data using Affymetrix microarrays, EGA, and TCGA for integrating clinical data and gene expression from PostgreSQL [18]. The standard p-value which is < 0.05 used to perform an analysis that moderate to the false or positive rate. The 5 hub genes for both up regulated and down regulated DEGs denoted with JetSet probe ID for analysed KM survival curves with expression split to low and high for comparison the expression value and recognized cut offs. KM plotter assessed between correlation and OS lung cancer.

The Hub Genes Expression Level

The GEPIA2 (<http://gepia2.cancer-pku.cn/#analysis>) was known analysis server of Gene Expression Profiling Interactive Analysis. It is utilized to analyzing the 9 736 tumours and 8 587 normal samples of RNA sequencing expression data [19]. The analyzing obtained using a standard processing pipeline from TCGA and GTEx ventures, and a box plot was used to determine the DEGs in various forms of tumor studies, which about the abnormal tissues and normal tissue [20]. The expression of the hub gene lung cancer for LUSC and LUAD was then confirmed. The data in visualisation connection of signature score was provided by the box plot. The features of GEPIA2 evaluated the huge data collection between the TCGA and GTEx samples and assessed the expression data. It looked into a variety of cancer gene types and identified possible biomarkers for therapeutic targets.

3 Results

Data Accumulation and Identification of DEGs

From 808 search results, chosen two different studies that are GSE176348 and GSE85841 datasets that contained a study involving profiles for gene expression analysis of lung adenocarcinoma and adjacent non-tumour tissues. Two platforms used in the work which are GPL25759 platform (Agilent-085753 Arraystar Human Epitranscriptomic microarray) in GSE176348 and GPL20115 platform (Agilent-067406 Human CBC lncRNA + mRNA microarray) in GSE85841.

The differentially expressed genes were determined by the GEO2R tool between two different of tissues using the $\log_2FCI > 1.0$ and p-values < 0.05 cut-off values as the guided principal value. GEOquery, Bioconductor/R, and the Limma programme were used to identify the top 250 DEGs. $\log_2FC \geq 1.0$ and $\log_2FC \leq -1.0$ were used to calculate

Table 1. Patients' information in GSE176348 and GSE85841 derived from the GEO database.

| Sample | Gross Sample | Type |
|-----------|--------------|---|
| GSE176348 | 24 | Non-small cell lung cancer and normal tissues |
| GSE85841 | 16 | Non-small cell lung cancer and normal tissues |

the identification DEGs for upregulated and downregulated genes, respectively. Among GSE176348 and GSE85841 obtained 88 genes upregulated and 183 genes downregulated of non-small cell lung cancer and adjacent non-tumour tissues. The GSE176348 dataset included 24 samples and dataset of GSE85841 with 16 samples of non-small cell lung cancer and adjacent non-tumour tissues. The two datasets of expression profile present the information of patients as displayed in Table 1. The retrieval of volcano plot of tumour vs normal along with mean-difference plot tumour vs normal for both the GEO accession has been showed in Fig. 1a, 1b, 2a and 2b. The number of common genes of upregulated and respective down regulated were displayed at the overlapping of area of the Venn diagram in Fig. 3.

Functional Enrichment Analysis

The DEGs were extracted using the functional annotation created by DAVID. The terms and pathway of GO enriched the suitable term for upregulated and downregulated biological process (BP), molecular function (MF) and cellular component (CC). The dataset generated a GO biological process. The modified Fisher precise EASE score of 0.05 with a p-value < 0.05 and an FDR value < 0.05 which was considered to be a strong enrichment value.

The process GO biological of upregulated arrived to the negative regulation of transcription from RNA polymerase II promoter and positive regulation of transcription from RNA polymerase II promoter. For GO molecular function for upregulated attained zinc ion binding and protein binding while cellular component presented nucleoplasm, integral component of membrane, cytosol and extracellular exosome.

The DEGs of biological process downregulated included G-protein coupled receptor signalling pathway and positive regulation of transcription from RNA polymerase II promoter. Besides, the molecular function of GO gained zinc ion binding, calcium ion binding and protein binding. Then, GO cellular component displayed plasma membrane, integral component of membrane, extracellular exosome and cytosol. These results are shown in Table 2. The KEGG pathway upregulated is metabolic pathways. The KEGG pathway of downregulated DEGs are metabolic pathways and pathways in cancer. The pathways are represented in Table 3.

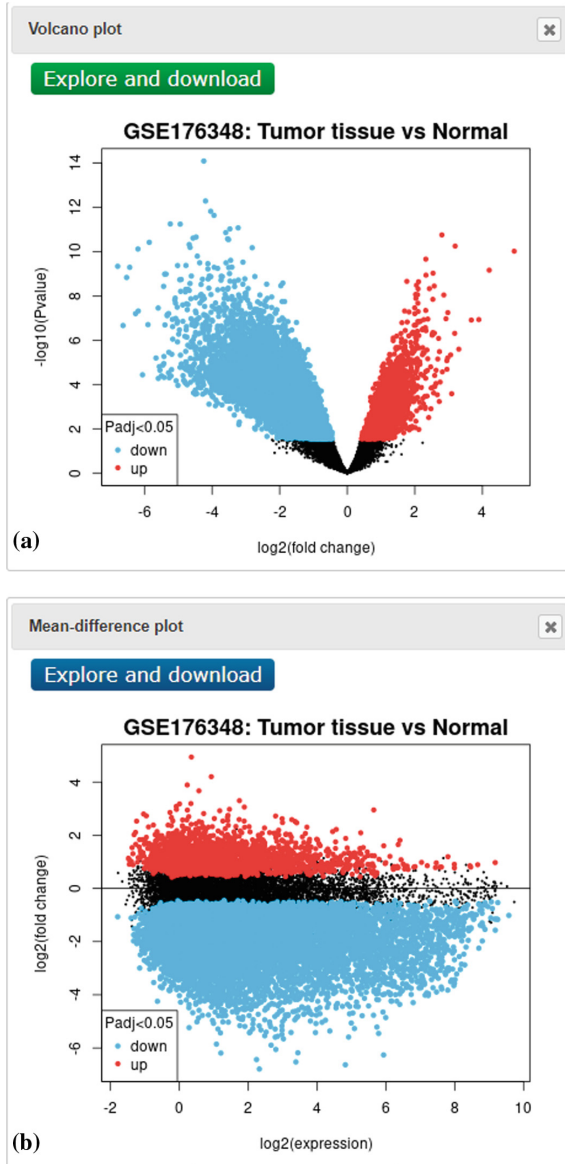


Fig. 1. a. Volcano plot of GSE176348 tumour tissues vs normal. b. Mean-difference plot of GSE176348 tumour tissues vs normal.

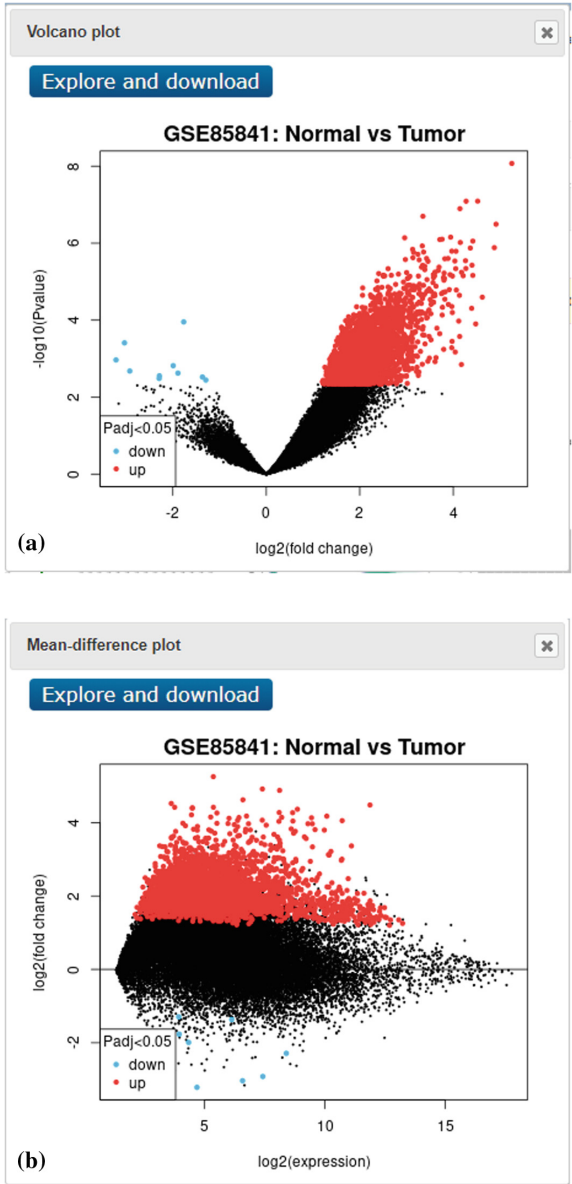


Fig. 2. a. Volcano plot of GSE85841 tumour tissues vs normal. b. Mean-difference plot of GSE85841 tumour tissues vs normal

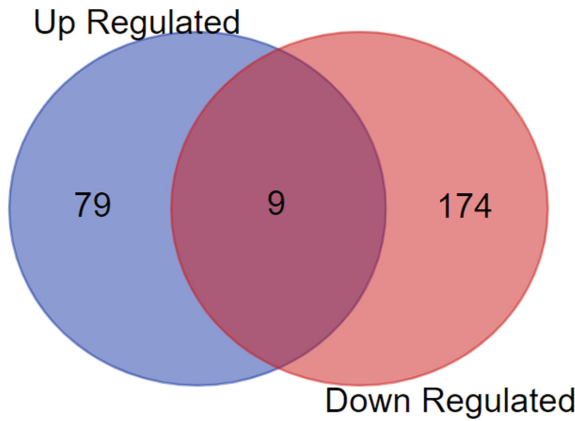


Fig. 3. Venn diagram of up regulated and down regulated genes.

Table 2. Gene Ontology (GO) terms such as biological process (BP), molecular functions (MF), and cellular component (CC) of DEGs by using DAVID v6.8 online tools.

| Expression | Category | Term | count | p-value | FDR |
|------------------|------------------|--|-------|---------|--------|
| Down - regulated | GOTERM_BP_DIRECT | G-protein coupled receptor signalling pathway | 12 | 1.6E-1 | 1.0E0 |
| | GOTERM_BP_DIRECT | Positive regulation of transcription from RNA polymerase II promoter | 17 | 1.2E-2 | 1.0E0 |
| | GOTERM_MF_DIRECT | Zinc ion binding | 16 | 6.7E-2 | 1.0E0 |
| | GOTERM_MF_DIRECT | Calcium ion binding | 16 | 1.0E-3 | 6.9E-2 |
| | GOTERM_MF_DIRECT | Protein binding | 87 | 2.5E-2 | 1.0E0 |
| | GOTERM_CC_DIRECT | Plasma membrane | 51 | 4.7E-3 | 1.3E-1 |
| | GOTERM_CC_DIRECT | Integral component of membrane | 54 | 7.8E-2 | 1.0E0 |
| | GOTERM_CC_DIRECT | Extracellular exosome | 36 | 1.4E-2 | 2.8E-1 |
| | GOTERM_CC_DIRECT | Cytosol | 38 | 5.3E-2 | 8.0E-1 |
| Up - regulated | GOTERM_BP_DIRECT | Negative regulation of transcription from RNA polymerase II promoter | 6 | 1.9E-1 | 1.0E0 |
| | GOTERM_BP_DIRECT | Positive regulation of transcription from RNA polymerase II promoter | 7 | 2.3E-1 | 1.0E0 |

(continued)

Table 2. (continued)

| Expression | Category | Term | count | p-value | FDR |
|------------|------------------|--------------------------------|-------|---------|--------|
| | GOTERM_MF_DIRECT | Zinc ion binding | 8 | 2.2E-1 | 1.0E0 |
| | GOTERM_MF_DIRECT | Protein binding | 50 | 2.7E-3 | 1.6E-1 |
| | GOTERM_CC_DIRECT | Nucleoplasm | 13 | 5.0E-1 | 1.0E0 |
| | GOTERM_CC_DIRECT | Integral component of membrane | 14 | 9.9E-1 | 1.0E0 |
| | GOTERM_CC_DIRECT | Cytosol | 27 | 6.2E-4 | 2.6E-2 |
| | GOTERM_CC_DIRECT | Extracellular exosome | 20 | 2.0E-2 | 5.4E-1 |

Table 3. KEGG pathway analysis of differentially expressed genes associated with lung cancer by using DAVID v6.8 online tools.

| Expression | Term | count | p-value | FDR |
|----------------|--------------------|-------|---------|--------|
| Down-regulated | Metabolic pathways | 8 | 9.7E-1 | 1.0E0 |
| | Pathways in cancer | 10 | 1.6E-2 | 9.5E-1 |
| Up-regulated | Metabolic pathways | 7 | 6.3E-1 | 1.0E0 |

Protein-protein interaction (PPI) network and Hub genes identification

The PPI of DEGs downregulated and upregulated analyzed using the STRING server as shown in Fig. 4 and Fig. 5. The confidence score was fixed at cut off > 0.9 for nodes appoint the significance of the PPI network. The query protein constructed the interaction through Cytoscape v 3.9.0 as PPI network visualization. The network of upregulated was 82 nodes and 32 edges in Fig. 6 and downregulated, 169 nodes and 139 edges of protein-protein interaction network in Fig. 7. The nodes indicated the total of proteins and edges as the interaction between nodes.

Both hub genes DEGs analyzed by the cytoHubba tool and the results shows the top-ranked of protein by MCC method as shown in Table 4 and Table 5. For up regulated, the top 5 hub genes network contains 5 nodes with 6 edges while for downregulated, top 5 hub genes network contains 5 nodes with 10 edges. Figure 8 and 9 represents the PPI network of the top 5 ranked for upregulated and downregulated DEGs. By analysis PPI network, the top 5 hub genes upregulated were generated ADH1B, CAV1, GSTA1, ADH1C, and ADH1A. The genes were referred as alcohol dehydrogenase 1B (class I) (ADH1B), Caveolin 1 (CAV1), Glutathione S-Transferase Alpha 1 (GSTA1), Alcohol Dehydrogenase 1C (Class I) (ADH1C), and Alcohol Dehydrogenase 1A (Class I) (ADH1A). The downregulated DEGs of the top 5 hub genes obtained CXCL12, FGF1, PPARG, FGF2 and IL1A. The genes represent C-X-C motif chemokine ligand 12 (CXCL12), Fibroblast growth factor 1 (FGF1), Peroxisome proliferator activated receptor gamma (PPARG), Fibroblast growth factor 2 (FGF2) and Interleukin 1 alpha (IL1A).

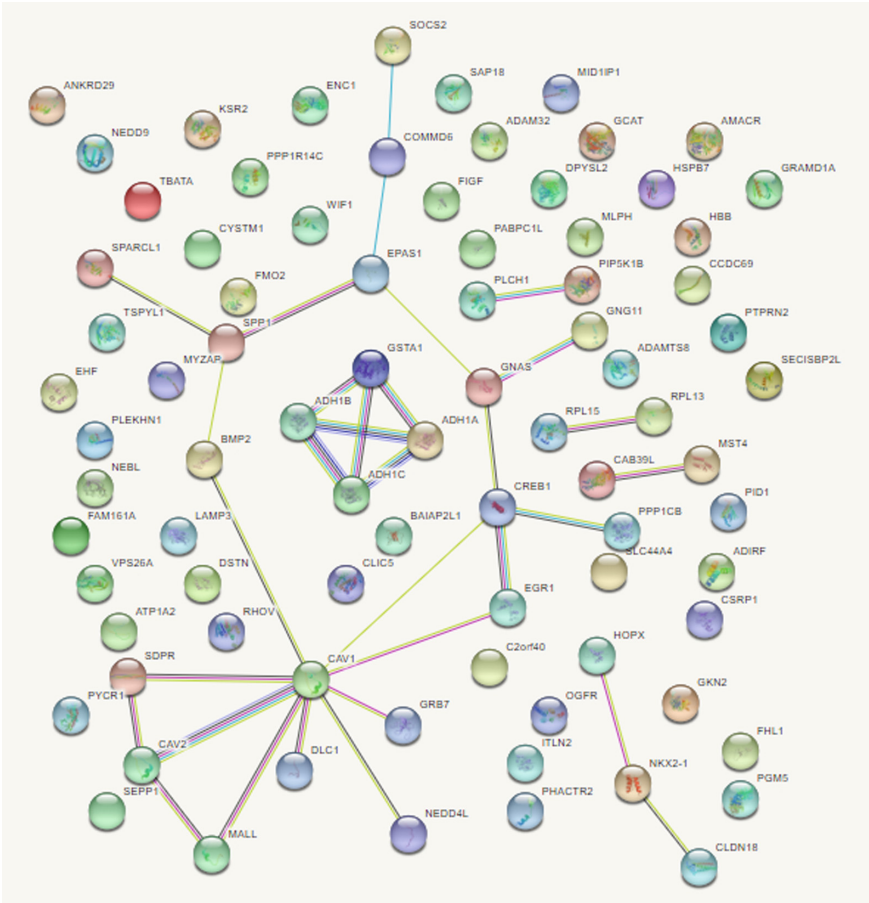


Fig. 4. Protein-protein interaction network of up regulated lung cancer with 82 nodes and 32 edges. Diagram retrieved from STRING database.

The Validation of Survival Analysis Hub Genes

The data of prognostic for both 5 hub genes generated the KM plotter online server to shows the prognostic value. The curves of results analyzed 40 patients lung cancer sample within a selected parameter. The upregulated DEGs of prognostic values presented the valid Affymetric ID 215766_at (GSTA1) was high expression that correlated with worse of overall survival in all patient's lung cancer with HR = 1.08, CI: 0.95–1.22, P = 0.24. The significantly high expression of worse OC was Affymetric ID 207820_at (ADH1A) with HR = 1.02, CI: 0.9–1.15, P = 0.8, Affymetric ID 212097_at (CAV1) with HR = 0.74, CI:0.65–0.84, P = 2.1e-06 and Affymetric ID 206262_at (ADH1C) with HR = 0.68, CI: 0.6–0.77, P = 2.7e-09 but not in Affymetric ID 209613_at (ADH1B) with HR = 0.67, CI:0.59–0.76, P = 5.3e-10. Figure 10 displayed the expression level hub genes.

The downregulated hub genes are displayed the expression survival of Affymetric ID 210118_s_at (IL1A) with HR = 1.41, CI: 1.24–1.6, P = 9.4e-08 where the high

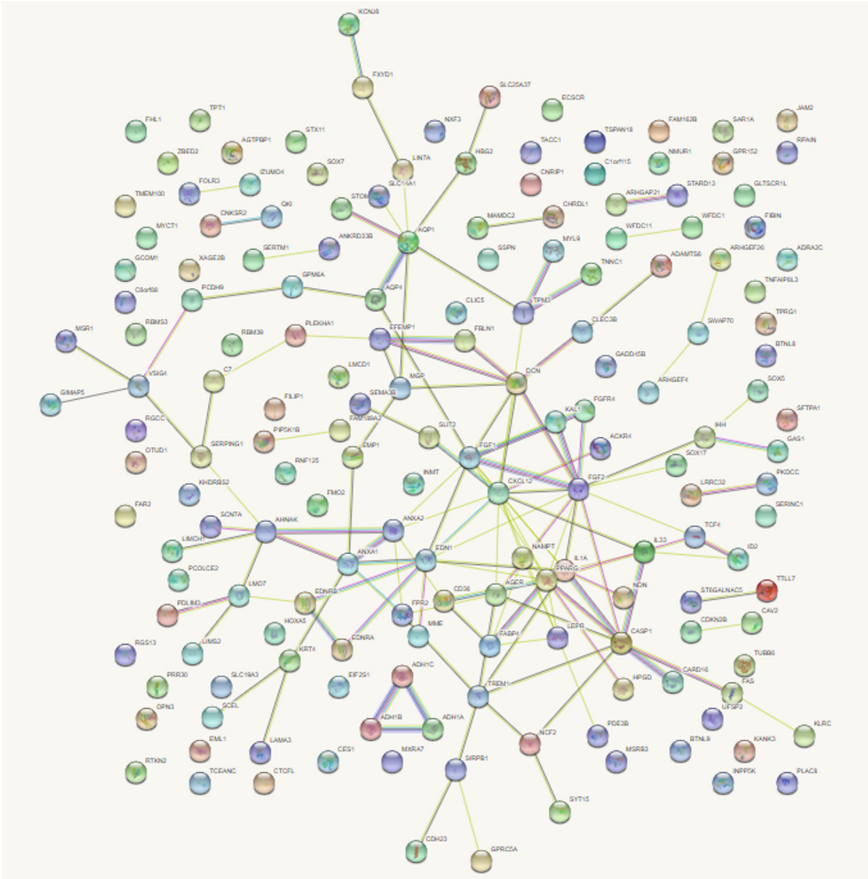


Fig. 5. Protein-protein interaction network of down regulated lung cancer with 169 nodes and 139. Diagram retrieved from STRING database

expression mRNA associated the worse overall survival for lung cancer. Affymetrix ID 208510_s_at (PPARG) with HR = 0.94, CI: 0.83–1.07, P = 0.37 is significant correlated shows the possibility survival lung cancer is worse. But it is not correlated to all OS lung cancer patients such as Affymetrix ID 203666_at (CXCL12) with HR = 0.93, CI: 0.82–1.06, P = 0.3 and Affymetrix ID of 204422_s_at (FGF2) with HR = 0.77, CI: 0.68–0.87, P = 3.9e-05. Affymetrix ID 205117_at (FGF1) mRNA expression portrayed the good OS with HR = 0.75, CI: 0.66–0.85 P = 8.4e-06. These were displayed in Fig. 11.

Expression Level of Hub Genes

In comparison to normal tissue samples, the expression level was evaluated box plot diagram by GEPIA2 tool verified both hub genes: downregulated and upregulated considerably over expressed in two tissue samples. The LUAD and LUSC of lung cancer with normal tissue samples are represented with red and blue boxes. A median is the

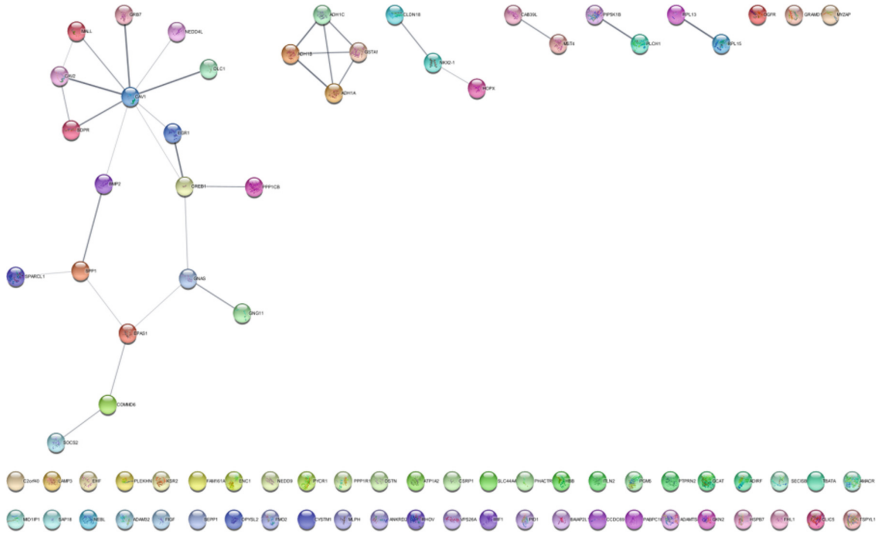


Fig. 6. Protein-protein interaction network of up regulated lung cancer with 82 nodes and 32 edges. Diagram retrieved from STRING in Cytoscape (V 3.9.0) software.

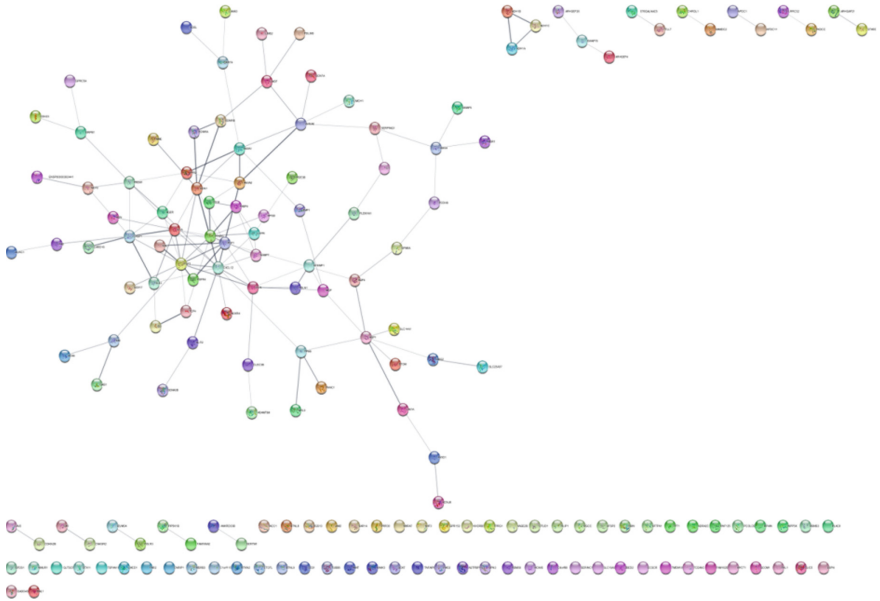


Fig. 7. Protein-protein interaction network of down regulated lung cancer with 169 nodes and 139. Diagram retrieved from STRING in Cytoscape (V 3.9.0) software.

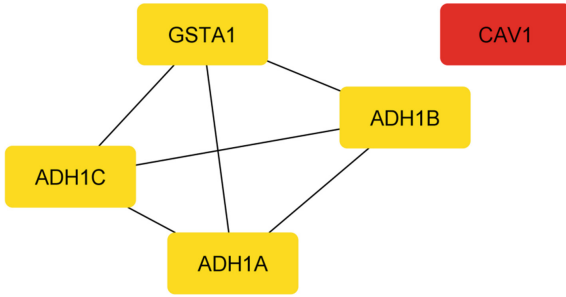


Fig. 8. Top 5 hub genes of up regulated of lung cancer. Diagram retrieved from cytoHubba plugin Cytoscape v 3.9.0.

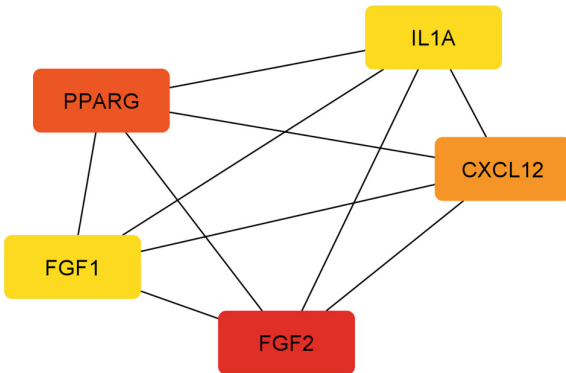


Fig. 9. Top 5 hub genes of down regulated of lung cancer. Diagram retrieved from cytoHubba plugin Cytoscape v 3.9.0.

Table 4. Up regulated top 5 hub genes network ranked by MCC method

| Top 5 hub genes network ranked by MCC method | | | |
|--|----------------------|--------------|-------|
| Rank | Shared Name | Display Name | Score |
| 1 | 9606.ENSP00000339191 | ADH1B | 10 |
| 2 | 9606.ENSP00000209668 | CAV1 | 6 |
| 2 | 9606.ENSP00000306606 | GSTA1 | 6 |
| 2 | 9606.ENSP00000335620 | ADH1C | 6 |
| 2 | 9606.ENSP00000426083 | ADH1A | 6 |

Table 5. Down regulated top 5 hub genes network ranked by MCC method

| Top 10 in network STRING network ranked by MCC method | | | |
|--|----------------------|---------------|-------|
| Rank | Shared Name | Display Name | Score |
| 1 | 9606.ENSP00000264498 | CXCL12 | 144 |
| 2 | 9606.ENSP00000287820 | FGF1 | 143 |
| 3 | 9606.ENSP00000379140 | PPARG | 141 |
| 4 | 9606.ENSP00000263339 | FGF2 | 133 |
| 4 | 9606.ENSP00000480791 | IL1A | 133 |

middle-positioned thick straight line. Each box's lower and upper boundaries represented by first and third quartile, respectively. Error bars in the bottom and top bars contains that shows the expression values of minimum and maximum data. Asterisk, which is statistically significant with every dot, denotes the tumour or normal tissues, respectively, for differential expression analysis using the ANOVA approach [21]. For up regulated, the dot indicator displayed in both LUAD and LUSC for ADH1B, CAV1 and GSTA1 while there is no dot indicator displayed in ADH1C and ADH1A as shown in Fig. 12. Thus, for down regulated, the dot indicator displayed in both LUAD and LUSC for CXCL12 and FGF2, while in PPARG the dot indicator only presents in LUSC and absence of dot indicator in FGF1 and IL1A as shown in Fig. 13.

4 Discussion

Previous studies on hub genes identification on lung cancer that was done by Sangeetha Subramaniam et al., provide a better understand that biomarker discovery for lung cancer has made significant progress where exhaled breath testing and circulating miRNA tests look to be particularly promising for early detection. Prognostic and therapeutic markers have thus far relied on the isolation of malignant tissue, whether from the lungs or circulating tumour cells [22]. Expression profiling and target genes of the lung cancer are important the biomarkers studies with the lung cancer pathogenesis. 250 total of DEGs was included further studies for gene expression profiles and attempted to predict top 5 hub genes each in both up regulated and down regulated that plays a role in lung carcinoma using bioinformatics approach.

We Furthermore, DEGs were identified in MF, BP, and CC lung cancers in order to obtain functional annotation as a result of GO term up regulated and down regulated. DEGs have the ability to produce extracellular vesicles from tumour cells as a result of LUSC and LUAD. The biological processes involved in G-protein coupled receptor signalling pathway, positive and negative regulation of transcription from RNA polymerase II promoter, while cellular component involved in plasma membrane, integral component of membrane, extracellular exosome, cytosol, nucleoplasm. Thus, molecular function involves in binding of zinc ion, calcium ion and protein. The both DEGs of KEGG pathway resulted affiliated with metabolic pathways and pathways in cancer. The PPI network was closely crossed by the top five (5) hub genes discovered distinguish

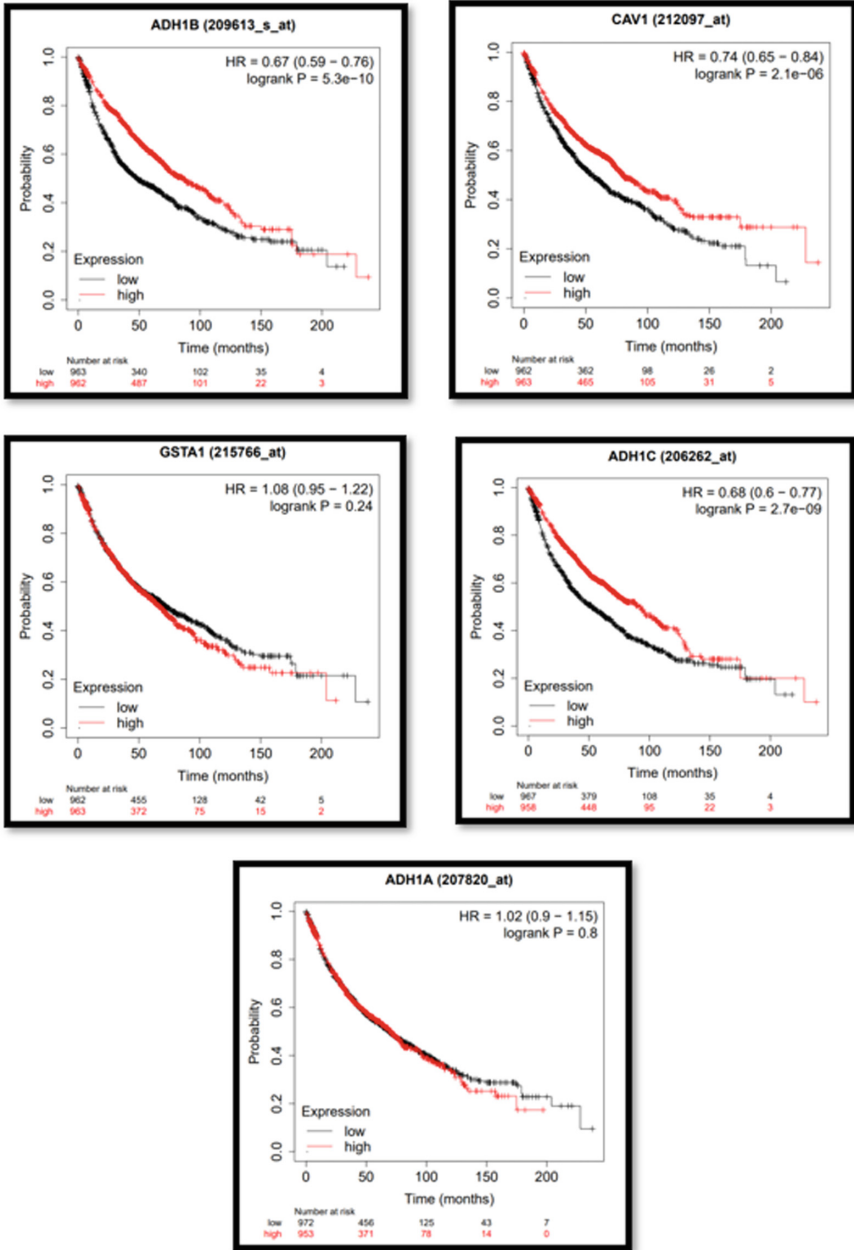


Fig. 10. Prognostic value of five genes with desired valid Affymetrix IDs are ADH1B (209613_at), CAV1 (212097_at), GSTA1 (215766_at), ADH1C (206262_at) and ADH1A (207820_at) in up regulated of lung cancer. HR: hazard ratio, CI: confidence interval. Diagram obtained from KM plotter online server.

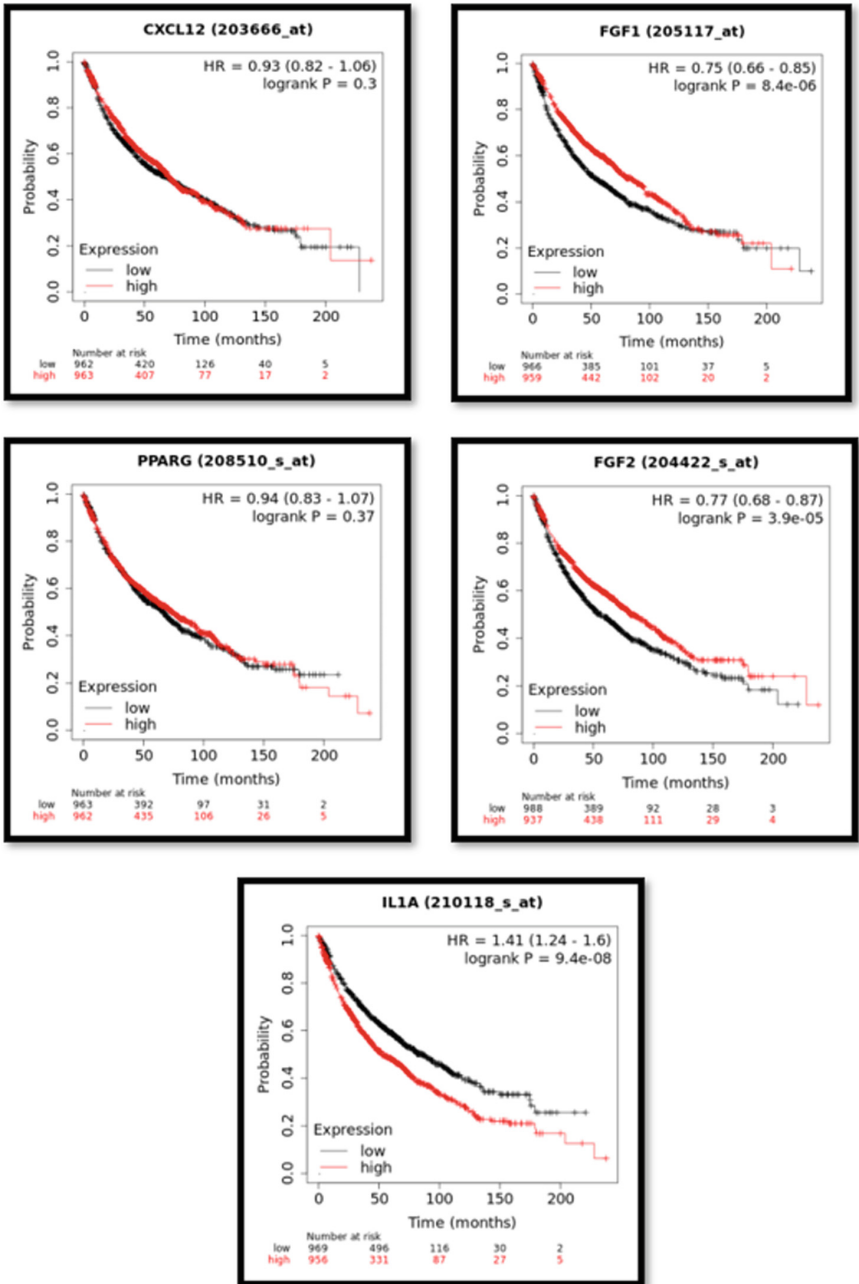
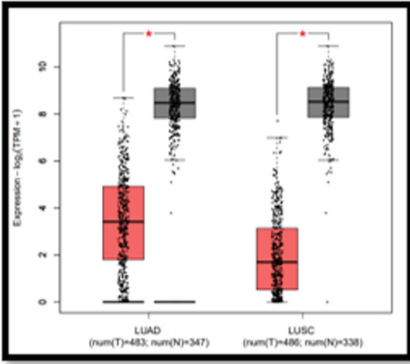
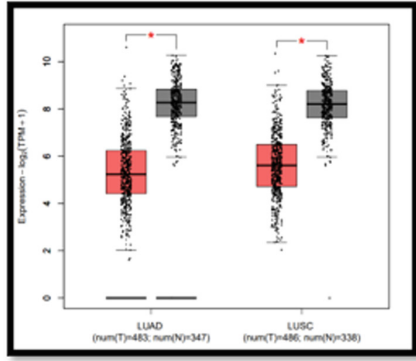


Fig. 11. Prognostic value of five genes with desired valid Affymetrix IDs are CXCL12 (203666_at), FGF1 (205117_at), PPARG (208510_s_at), FGF2 (204422_s_at) and IL1A (210118_s_at) in down regulated of lung cancer. HR: hazard ratio, CI: confidence interval. Diagram obtained from KM plotter online server.

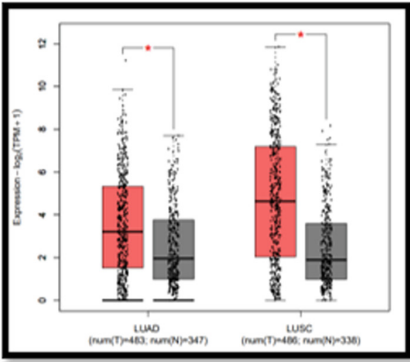
a. ADH1B



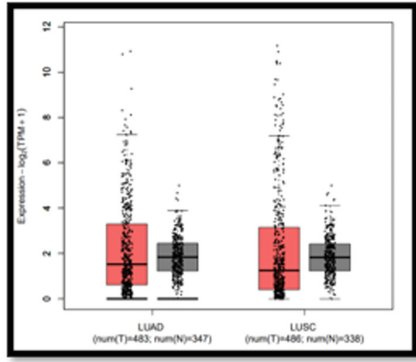
b. CAV1



c. GSTA1



d. ADH1C



e. ADH1A

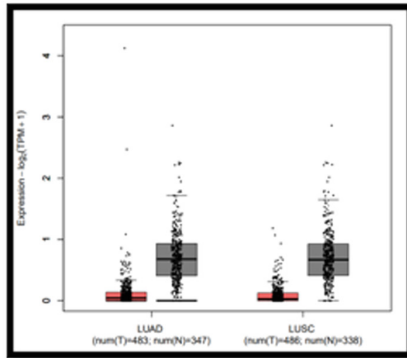
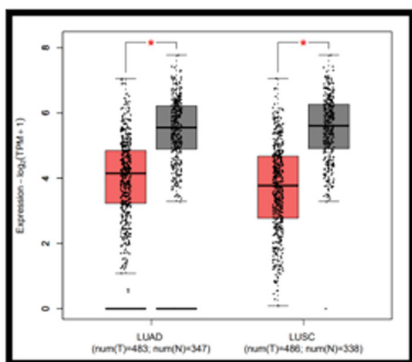
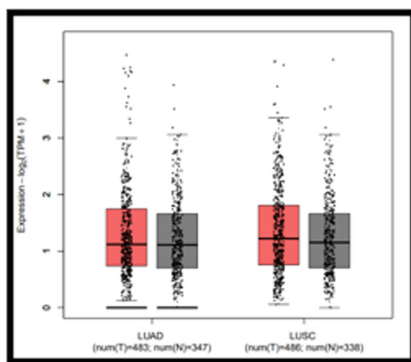


Fig. 12. Expression level of five hub genes of up regulated ADH1B, CAV1, GSTA1, ADH1C and ADH1A in up regulated of lung cancer.

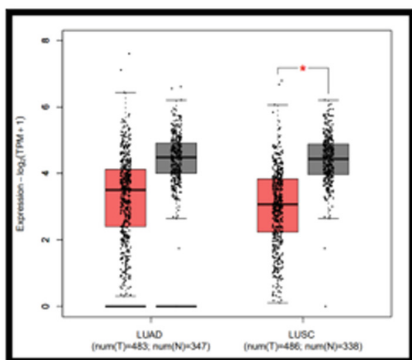
a. CXCL12



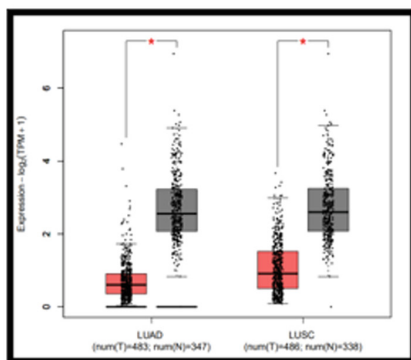
b. FGF1



c. PPARG



d. FGF2



e. IL1A

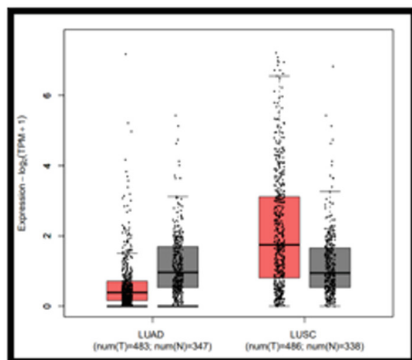


Fig. 13. Expression level of five hub genes of up regulated CXCL12, FGF1, PPARG, FGF2 and IL1A in down regulated of lung cancer.

with a high confidence level. It also finds that gene and interaction play a role in the development and progression of lung cancer, particularly in the diagnosis of NSCLC and SCLC. PPI network recognised and identified the top 5 hub genes.

Based on the studies, three genes which are ADH1A, ADH1B and ADH1C of up regulated are code for ADH enzymes that belongs to the alcohol dehydrogenase family. The expression of alcohol dehydrogenase (ADH) enzymes is linked to the prognosis of NSCLC and adenocarcinoma patients. Ethanol, retinol, other aliphatic alcohols, hydroxysteroids, and lipid peroxidation products are among the substrates metabolised by members of this enzyme family [10]. Ethanol metabolic processes have been discovered mostly in hepatic tissue, but also in the stomach, pancreas, lung, and brain [23]. Class 1 alcohol dehydrogenase has a high ethanol oxidation activity and is involved in ethanol catabolism. It is made up of multiple homo- and heterodimers of alpha, beta, and gamma subunits. A gene cluster is formed by the tandem organisation of three genes encoding alpha, beta, and gamma subunits in a genomic tract [24]. Thus, in ADH1B gene, it has two transcript variants that code for distinct isoforms that associate with lung cancer [25].

Caveolin-1 (Cav-1) is a major component of the plasma membrane's Ω -shaped caveolae. Cav-1 expression abnormalities in lung cancer have been connected to cancer progression, including cell proliferation, migration, apoptosis, and treatment resistance. As a result, a thorough understanding of the relationship between Cav-1 and lung cancer is required [26]. Cav-1 plays a context-dependent impact in lung cancer as well. Cav-1 expression in lung cancer is significantly lower than in normal pulmonary tissue, and its expression in cancer tissues of various histological kinds and stages also varies. Cav-1 expression in the parenchyma is higher in SCLC than NSCLC, and it is lower in advanced stages than early stages [27].

Glutathione S-transferase A 1 (GSTA1) is a GST isoform that is involved in the detoxification of electrophilic chemicals through glutathione conjugation. Earlier studies have shown that changing GST gene expression raises the risk of prostate cancer and hepatocellular carcinoma [28]. GSTA1 is a glutathione-binding enzyme that adds glutathione to electrophilic substances such as carcinogens, medicinal medicines, environmental pollutants, and oxidative stress products. This process is crucial in the detoxification of these chemicals. This gene located in the cytoplasm or in membrane of cells. Pan et al. have discovered GSTA1's promise in the early detection and treatment of lung cancer [29]. GSTA1 expression was shown to be higher in lung cancer tissues and cells than in healthy tissues and cells, and GSTA1 inhibition inhibited lung cancer cell proliferation [30].

CXCL12 is substantially expressed in primary lung cancer, as well as the bone marrow, liver, adrenal glands, and brain, all of which are potential locations for lung cancer metastasis. CXCL12 enhances cancer cell local invasion, whereas CXCL12 deficiency promotes tumour cell migration to tissues with high CXCL12 expression, such as the liver, bone marrow, and lung [31]. CXCL12's significance in NSCLC malignant proliferation is revealed by two lines of data. The first comes from retrospective investigations of human metastatic tumours, in which high CXCL12 expression is linked to a later stage of disease and a worse prognosis. The second piece of evidence comes from in vitro and in vivo studies that suggest CXCL12 interactions in the tumour microenvironment may

promote local tumour growth and that tumour cells that express high levels of CXCR4 have a higher invasive and metastatic potential [32].

Two FGF1 and FGF2 in down regulated, produces a member of the fibroblast growth factor (FGF) family of proteins. The oncogenic receptor tyrosine kinase (RTK) fibroblast growth factor receptor 1 (FGFR1) plays critical roles in physiological processes and cancer progression. Growth hormones, such as fibroblast growth factor 2 (FGF2), induce FGFR1 signalling, which results in receptor dimerization and transphosphorylation of FGFR1 [33]. FGF1 have a wide range of mitogenic and cell survival functions and FGF2 bind heparin and have a wide range of mitogenic and angiogenic properties. This protein has been linked to a number of biological processes, including tumour growth. FGF2 has been identified as the primary cause of chemo-resistance. FGF2 inhibits etoposide-induced cell apoptosis in SCLC cells via post-transcriptionally modulating inhibitor of apoptosis protein (IAP) levels and lowering the release of second mitochondria-derived activator of caspases (Smac) (progress caused cell death in response to cell stress) [34].

Followed by, PPARG gene The peroxisome proliferator-activated receptor (PPAR) subfamily of nuclear receptors is encoded by this gene. PPARs and retinoid x receptors (RXRs) form heterodimers, which regulate transcription of numerous genes [35]. Another class of nuclear hormone receptors, the peroxisome proliferator-activated receptors (PPARs), has long been recognized for its function in lipid and glucose metabolic control [1]. Other cellular functions, such as carcinogenesis-relevant cell differentiation, proliferation, survival, and apoptosis, are now known to be regulated by PPARs, with these effects mediated by transcriptional activation or repression of PPAR target genes [36].

Lastly, IL1A produces a protein that belongs to the interleukin 1 cytokine family. This cytokine is a pleiotropic cytokine that has a role in immunological responses, inflammatory processes, and hematopoiesis, among other things. Monocytes and macrophages create this cytokine as a proprotein, which is proteolytically digested and released in response to cell damage, causing apoptosis [37]. Lung cancer risk is linked to polymorphisms in the IL1A genes. When compared to their less metastatic counterparts, highly metastatic human lung cancer cell lines were shown to express higher levels of IL1A and were able to generate tumors with accelerated angiogenesis and lymph angiogenesis. The activation of angiogenesis and lymph angiogenesis appears to be IL1A mediated, as these effects can be reduced by an IL1R antagonist [38].

Researchers are analysing the relevance of gene biomarkers to improve the worsening situation of cancer diseases, with the development of integrative systems biology approaches. An analysis was carried out to create a PPI network comprising up-regulated and genes in order to show their connections and functional activities, with the top 5 genes being reported as the most important genes in each category. The KM plotter of lung cancer patient's database was used to assess the predictive values of individual genes, which revealed that chemokine ligand types are implicated in the relative receptor for cancer progression [39].

5 Conclusion

The top 250 DEGs in GSE176348, and GSE85841 of gene expression of lung cancer which are 88 upregulated and 183 downregulated. The mentioned information are integrated from the bioinformatics approach. The PPI network by Cytoscape and STRING consolidated the top 5 hub genes in lung cancer using Cytohubba tool in Cytoscape. The top 5 key genes upregulated were ADH1B, CAV1, GSTA1, ADH1C, and ADH1A while, downregulated were CXCL12, FGF1, PPARG, FGF2 and IL1A. The worsen OS of CAV1 and FGF2 part for prognostic of pathological with the inflammation process of tumour growth of cancer in lung. CXCL12 mRNA contains an high expression because of the good OS patients. These are useful for the understanding of biomarker hub genes in molecular biology of lung cancer and providing appropriate tools for prognosis prediction. It was also useful to gaining a deeper knowledge about the mechanism of lung cancer in order to determine risk, develop drugs, and find treatments.

Acknowledgement. The authors would like to acknowledge Department of Diagnostic and Allied Health Sciences, Faculty of Health and Life Sciences of Management and Science University, Shah Alam, Selangor Darul Ehsan, Malaysia for giving us opportunity and research facilities to conduct this research.

References

1. Ridge, C., McErlean, A. M. and Ginsberg, M. S. (2013) 'Epidemiology of Lung Cancer', *Seminars in Interventional Radiology*, 30(2), p. 93. doi: <https://doi.org/10.1055/S-0033-1342949>.
2. Sung, H. *et al.* (2021a) 'Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries', *CA: a cancer journal for clinicians*, 71(3), pp. 209–249. doi: <https://doi.org/10.3322/CAAC.21660>.
3. Pauk, N. *et al.* (2005) 'Lung cancer in women', *Lung Cancer*, 48(1), pp. 1–9. doi: <https://doi.org/10.1016/j.lungcan.2004.10.009>.
4. Sung, H. *et al.* (2021b) 'Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries', *CA: a cancer journal for clinicians*, 71(3), pp. 209–249. doi: <https://doi.org/10.3322/CAAC.21660z>.
5. Estimates, N. R. C. (US) P. on D. A. A. the A. of R. R. (1991) 'Cells of Origin for Lung Cancer'. Available at: <https://www.ncbi.nlm.nih.gov/books/NBK234231/> (Accessed: 24 December 2021).
6. Basumallik, N. and Agarwal, M. (2021) 'Small Cell Lung Cancer', *StatPearls*. Available at: <https://www.ncbi.nlm.nih.gov/books/NBK482458/> (Accessed: 24 December 2021).
7. Zappa, C. and Mousa, S. A. (2016) 'Non-small cell lung cancer: Current treatment and future advances', *Translational Lung Cancer Research*, 5(3), pp. 288–300. doi: <https://doi.org/10.21037/TLCR.2016.06.07>.
8. Chen, J. W. and Dhahbi, J. (2021) 'Lung adenocarcinoma and lung squamous cell carcinoma cancer classification, biomarker identification, and gene expression analysis using overlapping feature selection methods', *Scientific reports*, 11(1), p. 13323. doi: <https://doi.org/10.1038/S41598-021-92725-8>.

9. Barletta, J. A., Yeap, B. Y. and Chirieac, L. R. (2010) 'The Prognostic Significance of Grading in Lung Adenocarcinoma', *Cancer*, 116(3), p. 659. doi: <https://doi.org/10.1002/CNCR.24831>.
10. Lim, R. B. L. (2016) 'End-of-life care in patients with advanced lungcancer', *Therapeutic Advances in Respiratory Disease*, 10(5), p. 455. doi: <https://doi.org/10.1177/1753465816660925>.
11. Latimer, K. M. and Mott, T. F. (2015) 'Lung Cancer: Diagnosis, Treatment Principles, and Screening', *American Family Physician*, 91(4), pp. 250–256. Available at: <http://www.elsevier.com/locate/ymfs> (Accessed: 26 December 2021).
12. Clough, E. and Barrett, T. (2016) 'The Gene Expression Omnibus database', *Methods in molecular biology (Clifton, N.J.)*, 1418, p. 93. doi: https://doi.org/10.1007/978-1-4939-3578-9_5
13. Udhaya Kumar, S. *et al.* (2020) 'Analysis of differentially expressed genes and molecular pathways in familial hypercholesterolemia involved in atherosclerosis: A systematic and bioinformatics approach', *Frontiers in Genetics*, 11, pp. 1–16. doi: <https://doi.org/10.3389/FGENE.2020.00734/FULL>.
14. Luo, X. *et al.* (2019) 'High gene expression levels of VEGFA and CXCL8 in the peritumoral brain zone are associated with the recurrence of glioblastoma: A bioinformatics analysis', *Oncology Letters*, 18(6), pp. 6171–6179. doi: <https://doi.org/10.3892/OL.2019.10988>.
15. Kanehisa, M. and Goto, S. (2000) 'KEGG: Kyoto Encyclopedia of Genes and Genomes', *Nucleic Acids Research*, 28(1), p. 27. doi: <https://doi.org/10.1093/NAR/28.1.27>.
16. Doncheva, N. T. *et al.* (2019) 'Cytoscape StringApp: Network Analysis and Visualization of Proteomics Data', *Journal of Proteome Research*, 18(2), pp. 623–632. doi: <https://doi.org/10.1021/ACS.JPROTEOME.8B00702>.
17. Fehringer, G. *et al.* (2016) 'Cross-cancer genome-wide analysis of lung, ovary, breast, prostate, and colorectal cancer reveals novel pleiotropic associations', *Cancer Research*, 76(17), pp. 5103–5114. doi: <https://doi.org/10.1158/0008-5472.CAN-15-2980>.
18. Yang, M. *et al.* (2018) 'Differentially expressed and survival-related proteins of lung adenocarcinoma with bone metastasis', *Cancer Medicine*, 7(4), p. 1081. doi: <https://doi.org/10.1002/CAM4.1363>.
19. Tang, Z. *et al.* (2019) 'GEPIA2: an enhanced web server for large-scale expression profiling and interactive analysis', *Nucleic acids research*, 47(W1), pp. W556–W560. doi: <https://doi.org/10.1093/NAR/GKZ430>.
20. Yang, Z. Q., Liu, Y. J. and Zhou, X. L. (2020) 'An integrated microarray analysis reveals significant diagnostic and prognostic biomarkers in pancreatic cancer', *Medical Science Monitor*, 26. doi: <https://doi.org/10.12659/MSM.921769>.
21. Yu, L. *et al.* (2020) 'Identification of SPRR3 as a novel diagnostic/prognostic biomarker for oral squamous cell carcinoma via RNA sequencing and bioinformatic analyses', *PeerJ*, 2020(6). doi: <https://doi.org/10.7717/PEERJ.9393/SUPP-4>.
22. Subramaniam, S. *et al.* (2013) 'Lung cancer biomarkers: State of the art', *Journal of Carcinogenesis*, 12. doi: <https://doi.org/10.4103/1477-3163.107958>.
23. Cederbaum, A. I. (2012) 'ALCOHOL METABOLISM', *Clinics in liver disease*, 16(4), p. 667. doi: <https://doi.org/10.1016/J.CLD.2012.08.002>.
24. Yasunami, M. *et al.* (1990) 'The human class I alcohol dehydrogenase gene cluster: three genes are tandemly organized in an 80-kb-long segment of the genome', *Genomics*, 7(2), pp. 152–158. doi: [https://doi.org/10.1016/0888-7543\(90\)90535-3](https://doi.org/10.1016/0888-7543(90)90535-3).
25. Wang, P. *et al.* (2018) 'Distinct Prognostic Values of Alcohol Dehydrogenase Family Members for Non-Small Cell Lung Cancer', *Medical Science Monitor : International Medical Journal of Experimental and Clinical Research*, 24, p. 3578. doi: <https://doi.org/10.12659/MSM.910026>.

26. Shi, Y. B. *et al.* (2020) 'Multifaceted Roles of Caveolin-1 in Lung Cancer: A New Investigation Focused on Tumor Occurrence, Development and Therapy', *Cancers*, 12(2). doi: <https://doi.org/10.3390/CANCERS12020291>.
27. Chen, D. *et al.* (2014) 'Duplex value of caveolin-1 in non-small cell lung cancer: a meta analysis', *Familial cancer*, 13(3), pp. 449–457. doi: <https://doi.org/10.1007/S10689-014-9707-6>.
28. Lee, W. H. *et al.* (1994) 'Cytidine methylation of regulatory sequences near the pi-class glutathione S-transferase gene accompanies human prostatic carcinogenesis', *Proceedings of the National Academy of Sciences of the United States of America*, 91(24), pp. 11733–11737. doi: <https://doi.org/10.1073/PNAS.91.24.11733>.
29. Pan, X. D. *et al.* (2014) 'Expression and function of GSTA1 in lung cancer cells', *Asian Pacific journal of cancer prevention : APJCP*, 15(20), pp. 8631–8635. doi: <https://doi.org/10.7314/APJCP.2014.15.20.8631>.
30. Mlakar, V. *et al.* (2021) 'The analysis of GSTA1 promoter genetic and functional diversity of human populations', *Scientific reports*, 11(1). doi: <https://doi.org/10.1038/S41598-021-83996-2>.
31. Wald, O., Shapira, O. M. and Izhar, U. (2013) 'CXCR4/CXCL12 Axis in Non Small Cell Lung Cancer (NSCLC) Pathologic Roles and Therapeutic Potential', *Theranostics*, 3(1), p. 26. doi: <https://doi.org/10.7150/THNO.4922>.
32. Wagner, P. L. *et al.* (2009) 'CXCL12 and CXCR4 in adenocarcinoma of the lung: association with metastasis and survival', *The Journal of thoracic and cardiovascular surgery*, 137(3), pp. 615–621. doi: <https://doi.org/10.1016/J.JTCVS.2008.07.039>.
33. Arnedos, M. *et al.* (2013) 'Fibroblast growth factor receptor inhibitors as a cancer treatment: from a biologic rationale to medical perspectives', *Cancer Discov*, 3(3), pp. 264–279. doi: <https://doi.org/10.1158/2159-8290.cd-12-0362>.
34. Pardo, O. E. *et al.* (2003) 'Fibroblast growth factor 2-mediated translational control of IAPs blocks mitochondrial release of Smac/DIABLO and apoptosis in small cell lung cancer cells', *Molecular and cellular biology*, 23(21), pp. 7600–7610. doi: <https://doi.org/10.1128/MCB.23.21.7600-7610.2003>.
35. Tyagi, S. *et al.* (2011) 'The peroxisome proliferator-activated receptor: A family of nuclear receptors role in various diseases', *Journal of Advanced Pharmaceutical Technology & Research*, 2(4), p. 236. doi: <https://doi.org/10.4103/2231-4040.90879>.
36. Theocharis, S. *et al.* (2002) 'Expression of peroxisome proliferator activated receptor-gamma in non-small cell lung carcinoma: correlation with histological type and grade', *Lung Cancer*, 36(3), pp. 249–255. doi: [https://doi.org/10.1016/S0169-5002\(02\)00013-2](https://doi.org/10.1016/S0169-5002(02)00013-2).
37. Zhang, J. M. and An, J. (2007) 'Cytokines, Inflammation and Pain', *International anesthesiology clinics*, 45(2), p. 27. doi: <https://doi.org/10.1097/AIA.0B013E318034194E>.
38. Watari, K. *et al.* (2014) 'Tumor-Derived Interleukin-1 Promotes Lymphangiogenesis and Lymph Node Metastasis through M2-Type Macrophages', *PLoS ONE*, 9(6), p. 99568. doi: <https://doi.org/10.1371/JOURNAL.PONE.0099568>.
39. Zheng, Z. *et al.* (2018) 'CXCL13/CXCR5 Axis Predicts Poor Prognosis and Promotes Progression Through PI3K/AKT/mTOR Pathway in Clear Cell Renal Cell Carcinoma', *Frontiers in Oncology*, 8(JAN), p. 682. doi: <https://doi.org/10.3389/FONC.2018.00682>.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

