EXPLORING HUB GENES IN LUNG CANCER USING INTEGRATED BIOINFORMATICS ANALYSIS

Nur Azznira Azman, Venkataramanan Swaminathan*

*Department Diagnostic and Allied Health Science, Faculty of Health Sciences, Management and Science University, University Drive, Off Persiaran Olahraga, Section 13, 40100 Shah Alam, Selangor Darul Ehsan, Malaysia

Corresponding author: s_ venkataramanan @msu.edu.my

ABSTRACT

Lung cancer is a serious health issue worldwide causing death men than women. The environmental and genetic factored in metastasis primary lung cancer. The survival rate patient improved by identification hub genes lung cancer with bioinformatics tools. The present study was to determine the biological pathway and PPI network by identification the biomarker hub genes lung cancer. The GSE84797, GSE28827, and GSE115456 were retrieved from GEO attained upregulated and downregulated by GEO2R tool. Enrichment analysis DEGs analyzed used DAVID server. PPI network and hub genes was constructed by STRING, Cytoscape and cytoHubba. The OS and expression level were retrieved by KM plotter server and GEPIA2. Results obtained 3 DEGs dataset by GEO2R with -1.0 ≤ logFC≥1.0.Upregulated and downregulated DEGs presence for GO and pathway enrichment analysis.The PPI network and hub genes identification done with cutoff >0.9 obtained KRT5, KRT13, KRT17, KRTI6 KRTI5 , C5, PPBP, CXCL2, CCR9, and CCR7 .The OS done analyzed showed hazard ratio while expression level presented the different gene of LUSC/LUAD and normal tissues.The hub genes help in understanding the system biology of cancer through the identification network of protein interaction and hub gene identification lung cancer for classified the medicine and early diagnosis cancer.

Keywords: Primary lung Cancer, PPI network analysis, DEGs, Overall Survival analysis, Hub genes

INTRODUCTION

Lung cancer is a common cancer cause of death around the world with an estimated 13.3 million death in 2018 established on WHO and most death-causing cancer per year for both men and women in Malaysia. During 2018, the new cases rates for men (16.6%) higher than women (5.4%) and the leading top two causes of death of cancer after breast cancer in Malaysia. Lung cancer has been prevalence rising related to tobacco smoking (90%) cause lung cancer and 1-2% by outdoor air pollution and non-smoker [1].

Lung cancer arises from the cells of the respiratory epithelium. It can divide into two type's classes as well as small cell lung cancer (SCLC) and non-small cell lung cancer (NSCLC). Moreover, there are a group of histological subtypes of NSCLC patients allocated into three major pathologic subtypes as lung adenocarcinoma (LUAD), large cell carcinoma, and lung squamous cell carcinoma

(LUSC). Small cell lung cancer (SCLC) is related to smoking. Its connotation is resilient with lung squamous cell carcinoma (LUSC) and small cell lung cancer (SCLC) than lung adenocarcinoma (LUAD) [2-3]. An approximately of NSCLC at least 85% cases due from smoking whereas the rest 15% cases by SCLC [3].

Individual exposure to lung cancer may depend on interactions of gene enzymes affected by environmental, genetic history, and the consumption of tobacco. Thus, this exposure may affect the activation of pro-carcinogens and the level of formation addition DNA based on DNA lesion repairing mechanisms that transform to the pathogenesis of tumor suppressor genes. Lung cancer not only affected smokers but non-smoker will get risk.The primary lung cancer highly heterogeneous that occurs in the bronchial tree in different locations of the airways compartment. Besides, these variable symptoms and signs depending on anatomical location prognosis of the tumor [4] as detection of giving suitable treatments for patients.

The mutation of the clonal fraction is associated with a higher likelihood to patients with LUAD, then indicating there a greater tendency in early metastases during the development of a tumor. This study was presented to identify the cellular component, biological process, and molecular function of gene expression and determine the PPI network of lung cancer However, identification biomarker of hub genes helps investigate the overall survival and the occurrence of expression the lung cancer. The different subtypes of tumor rise in the origin of cells localized to the microenvironment which helps the identification of the complication of cancer biology of lung cancer. Thus, the bioinformatics tools are significant of high throughput sequencing will increase the analysis of the early assessment, staging, and prognosis likelihood [5].

METHODOLOGY

Gene Expression Data Processing

The data of expression profiling was collected from the GEO database in NCBI that was performed to access GSE115456, GSE28827, and GSE84797. The GEO dataset provided various experiments that able to download the expression profile. The platform of GSE115456, GSE28827, and GSE84797 created by GLP570 platform (Affymetrix Human Genome U133 Plus 2.0 Array) , GPL6947 platform (Illumina HumanHT-12 V3.0 expression bead chip) and GLP19965 (nCounter PanCancer Immune Profiling Panel). The three datasets were matched by histology as adenocarcinoma and squamous cell carcinoma that downloaded from a public database.

Data Processing and DEGs Selection

The raw gene expression data have been conducted by GEO2R tool that incorporated with GEOquery, Bioconductor/R and Limma package used for analysis the differential GEO data expression through table ordered of the gene with GEO Profiles graphic visualization [6].GEOquery, call parameter the file annotation of selected the distribution expression samples that finding suitable data sample analysis as data input. Limma extracted the top-ranked of genes for output data the top 250 genes through JSON formatting and HTML tables [6]. The identification of

DEGs using GEO2R tool matched and compared the cancer tissue in between adenocarcinoma and squamous cell carcinoma. The t-test and Benjamin-Hochberg were used for calculated p-value and FDR in the GEO2R tool. The DEGs screened accordingly principal standard $p<0.05$ and logFC>1 from the dataset. The upregulated and downregulated DEGs are considered as logFC ≥ 1 and logFC ≤ 1.

Functional Enrichment Analysis

The functional annotation associated Gene Ontology function and KEGG pathway enrichment by DAVID servers. The GO analysis the functional of genomic or transcription data on a large scale, although KEGG pathways analyse the data concerning the process of molecules or genes integration network [7]. The DAVID v6.8 investigated and identified the biological information base and analyse tool for extract the biological phenomena study of protein or gene list [8]. The study DEGs of enrichment the pathways, molecular and biological function GO with adjusted significant value (*p*-values<0.05 and FDR<0.05).

Construction of PPI Network and Hub Genes Identification

The PPI network is exploited by STRING online database for evaluation between the DEGs from the dataset's results. We fixed cut off>0.9 used as high confident interaction score to elude inaccurate of PPI network as strong accomplished for PPI network. The cut off eliminated the PPI network consistently. The PPI network was visualized using Cytoscape software v3.8.1 that was acquired from the STRING database. The STRING database with Cytoscape software as input data provided the high-level data with displayed the framework of enrichment network and functional detection [9] .The hub gene constructed from cytoHubba as finding the top 5 of hub genes upregulated and downregulated of DEGs that extracts the sub-network whereas the MCC was best performance for constructing PPI network [10].

The Validation of Survival Analysis Hub Genes

The Kaplan-Meier plotter was accomplished to evaluate the effect of 54 000 genes the mRNA, miRNA, and protein on survival based on 21 types of cancer including breast cancer, ovarian cancer, lung cancer, and gastric cancer. These used 13 316 cancer samples including 6 234 samples of breast cancer, 2190 samples of ovarian cancer, 3 452 samples of lung cancer, and 1 440 samples of gastric cancer. The hazard ratio (HR) with a 95% confidence interval and log-rank *P* have calculated the plot using the R and Bioconductor package [11] as showed the survival plot. The overall survival (OS) and relapse-free data were identified by the GEO database with Affymetrix microarrays, EGA, and TCGA for integrated the clinical data and gene expression from PostgreSQL. The analysis was performed with the standard *p*-value <0.05 that significantly moderate false and positive rate. The data of survival composed of 1 925 lung cancer samples compared gender, stage, grade, histology, age, and chemotherapy [12]. The 5 hub genes 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 was acknowledged as a web server of Gene Expression Profiling Interactive Analysis designed in analysed the 9 736 tumors and 8 587 normal samples of RNA sequencing expression data. It was attained from TCGA and GTEx ventures by a standard processing pipeline and determined the DEGs in different types of cancer analysis as well as abnormal tissues and normal tissue with a box plot. Then, validated the expression of hub gene lung cancer for LUSC and LUAD. The box plot provided the data in the visualization relationship of the signature score. The features of GEPIA2 assessed the expression data and compared the large data collection between TCGA and GTEx sample. It investigated the multiple cancer gene types and provided determination of potential biomarker in therapeutic targets [13].

RESULTS

Data Collection and Identification of DEGs

The GSE84797, GSE28827, and GSE115456 datasets contained profiles for gene expression in the LUAD and LUSC groups. A study in patients with primary lung cancer and non-small lung cancer carcinoma was obtained from the GEO database. There are three platforms used in this work such as GLP570 platform (Affymetrix Human Genome U133 Plus 2.0 Array), GPL6947 platform (Illumina HumanHT-12 V3.0 expression beadchip), and GLP19965 (nCounter PanCancer Immune Profiling Panel).

The DEGs were identified by the GEO2R tool between adenocarcinoma and squamous cell carcinoma using the $\log2FC| > 1.0$ and p-values < 0.05 cutoff values as the standard principal value. It was designed by GEOquery, Bioconductor/R, and Limma package for identified the top250 DEGs. The identification DEGs for upregulated and downregulated were calculated based on $\log FC \geq 1.0$ and $\log FC \leq 1.0$.Among GSE84797, GSE28827 and GSE115456 obtained 365 genes upregulated and 288 genes downregulated of adenocarcinoma and squamous cell carcinoma tissues. The GSE84797 dataset included 28 patients with primary lung cancer .Then dataset of GSE28827 and GSE115456 got 26 patients and 30 patients of NSLC, respectively. The three datasets of expression profile present the information of patients as shown in **Table 1**.

Functional Enrichment Analysis

The functional annotation constructed used DAVID serves intended to extract the DEGs. The terms and pathway of GO enriched the approximate term for upregulated and downregulated cellular component (CC), biological process (BP), and molecular function (MF). GO biological process from the dataset. 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 GO biological process of upregulated enriched the response to lipopolysaccharide, epidermis development, intermediate filament cytoskeleton organization, cytoskeleton organization, innate immune response. For GO cellular component for upregulated attained extracellular exosome and intermediate filament while molecular function presented structural constituent of the cytoskeleton, scaffold protein binding, protein binding, and structural molecule activity.

The DEGs of biological process downregulated included immune response, inflammatory response, regulations of complement activation, chemotaxis, and chemokine-mediated signalling pathway. Besides, the cellular component of GO gained plasma membrane, extracellular space, cell surface, extracellular region, and extracellular exosome. Then, GO molecular function displayed chemokine activity, receptor binding, and protein binding. These are shown in **Table 2**.The KEGG pathway of downregulated DEGs gained cytokine-cytokine receptor interaction, herpes simplex infection, the intestinal immune network for IgA production, complement and coagulation cascades, chemokine signalling pathway, and pertussis. The upregulated pathway is focal adhesion, leishmaniasis, and small cell lung cancer. The pathways are represented in **Table 3**.

PPI Network and Hub Genes Identification

The PPI of DEGs downregulated and upregulated evaluated used the STRING tool. The confidence score was set of cut off >0.9 for nodes designate the significance of the PPI network. The query protein constructed the interaction through Cytoscape v3.8.1 as PPI network visualization. The network of upregulated was 338 nodes and 212 edges in **Figure 1** and downregulated, 277 nodes and 276 edges of protein-protein interaction network in **Figure 2**. The nodes indicated the total of proteins and edges as the interaction between nodes. Both hub genes DEGs are obtained by the cytoHubba tool to generate the top-ranked of protein.

The top 5 hub genes network attained 5 nodes with 10 edges. **Figure 3** represents the PPI network of the top 5 ranked for upregulated and downregulated DEGs, respectively. By analysis of the PPI network, the top 5 hub genes upregulated were generated KRT5, KRT15, KRT13, KRT17, and KRT16. Among them, the genes were Keratin in type I and type II. The keratin of type cytoskeletal indicate type I cytoskeletal 5 (KRT5), type I cytoskeletal 15 (KRT15), type I cytoskeletal 13 (KRT13), type I cytoskeletal 17 (KRT17), and type I cytoskeletal 16 (KRT16).The downregulated DEGs of the top 5 hub genes obtained PPBP, C5, CXCL2, CCR7, and CCR9. The gene of PPBP was pro-platelet of basic protein in chemokine ligand 7 and C5 as C3 and PZP-like-alpha-2-macroglobulin domain that contain the protein 4. The CXCL2 and CCR7 act as macrophage inflammatory protein 2-alpha and Epstein-Barr virus-induced the G protein with coupled receptor 1 while CCR9, chemokine receptor 9.

The KM Plotter and Expression Level of Hub Genes

The data of prognostic for 5 hub genes upregulated and downregulated generated the KM plotter online server for determined the prognostic value. The survival curves are analyzed 1925 patients sample lung cancer using a selected parameter. The downregulated hub genes are displayed the expression survival of Affymetrix ID 207445_s_at, CCR9 (HR=1.14, 95% CI: 1-1.29, *P*= 0.0446) which high expression mRNA associated the worse overall survival for lung cancer (**Figure 4D)**. However, it not significantly correlated to all OS lung cancer patients such as CXCL2 (HR=0.88, 95% CI: 0.77-1, *P*= 0.0424) and CCR7 (HR=0.8, 95% CI: 0.7-0.9, *P*= 0.0004) in **Figure 4C** and **Figure 4E**. The valid Affymetrix ID of 2099771_x_at (C5) mRNA expression showed the good OS with HR=0.63, 95% CI: 0.55-0.71, *P*= 4.6e-

13 in **Figure 4A**. PPBP with Affymetrix ID 214146_s_at was significant correlated with worse overall survival lung cancer patients (HR=1.04, 95% CI: 0.91-1.17 *P*= 0.5904) in **Figure 4B**. These were displayed in **Figure 4**.

The upregulated of prognostic values presented the valid Affymetric ID 209800_at (KRT16) was high expression that correlated with worse of overall survival in all patients lung cancer (HR=1.55, 95% CI: 1.36-1.76, *P*= 1.5e-11) in **Figure 5D**. The significantly high expression of worse OC were KRT5 ,HR=1.27, 95% CI: 1.12-1.44, *P*=0.0002 **(Figure 5A)**, KRT13 , HR=1.17, 95% CI:1.03-1.32, *P*= 0.0161 **(Figure 5B)** and KRT17 ,HR=1.44, 95% CI: 1.27-1.64, *P*=1.3e-8 **(Figure 5E)** but not in KRT15 (HR = 0.97, 95% CI:0.85-1.1, *P*= 0.6059) in **Figure 5C**. **Figure 5** displayed the expression level hub genes.

The expression level was analysed box plot by GEPIA2 validated downregulated and upregulated hub genes significantly overexpressed in LUAD and LUSC tissue samples with comparison the normal tissue samples. The red and blues boxes represent the LUAD and LUSC of lung cancer with normal tissue samples. The thick straight line in the middle represents a median. Lower and upper limits each box for showed the first and third quartiles in correspondingly. The error bars in the bottom and top bars displayed as the minimum and maximum values expression data. The asterisk, significantly statistically with each dot indicates as a tumor or normal tissues, respectively used ANOVA method for differential expression analysis [14]. The C5 (LUSC) showed dot indicator only while PPBP and CXCL displayed in both LUSC and LUAD **(Figure 6)**. KRT5 (LUSC) gene only showed dot indicator while KRT13, KRT15, KRT16, and KRT17 appeared dot indicator for LUSC and LUAD **(Figure 7)**.

DISCUSSION

The development and incidence of lung cancer similar for each multi-stage of the development process such as genes and factors. Besides, in-depth understanding helps developed the effectiveness of therapeutic therapy and signs for early diagnosis of lung cancer [15]. Lung cancer expression profiles and target genes important in study of biomarkers with lung cancer pathogenesis .A total of 751 DEGs as considered further studies for gene expression profiles.

Besides, DEGs identification explored the MF, BP, and CC lung cancer for attaining the functional annotation to result of GO term upregulated and downregulated. The DEGs have the potential for tumor cell secreted the extracellular vesicle consequent of LUSC and LUAD. The cellular component and biological processes involved in protein regulation localized, cancer, and the injury of organismal. The both DEGs of KEGG pathway resulted associated with apoptosis, proliferation, and angiogenesis the gene product of LUAD and LUSC. Top five (5) hub genes found eminent with the high confidence score closely intersected the PPI network. Besides, it conclude to association of pathogenesis and progression lung cancer based on gene and interaction especially diagnosis NSCLC and SCLC. The top 5 hub genes recognized and identified by PPI network.

Keratin has expressed the adenocarcinoma and squamous cell carcinoma in lung cancer. KRT5 expressed the endometrial adenocarcinoma in stratified epithelial as an indicator of the squamous metaplasia [16]. It highly expression associated with

lymphatic metastasis and invasion of lymph nodal and vascular [17]. The marker of ancestor cell found in squamous of basal carcinoma subtypes with histology structures. KRT13 expressed the megakaryocytes of cancer cells for gained entry in bone or soft tissues with function dependent cell cancer types [18].The KRT15 redevelop the new hair follicles and maintain the follicle cell, it proven after adult mouse K15 promoter developed the hair follicle lump cell [19]. Thus, it localized in a stem cell as feature of the lump. The normal epithelium stratified KRT5 and KRT 15 expression restricted to the proliferation of basal keratinocytes and precise the differentiation cell to develop a new cell. So KRT15 activated the FOXM1 for inducing stem cell expansion as expressed outer the root sheath in the epidermis of the hair follicle.

KRT17 played the role for biological process and cell division whereas perceived in the various condition of tumour progression which AKT or mTOR pathway involved in normal and cancerous. It works as growth regulation for contributed to cell cycle progression, prognosis, proliferation, and tumor development [20-21] . The phosphorylation activated the AKT in Thy308, then Ser473 added EMT for migrates the activity cancer exacerbation and increases the mesenchymal properties [21-22].It was a marker of basal cell differentiation in complex epithelia and analytical of a certain type of epithelial 'stem cells' that acts as a promoter of epithelial proliferation. It also act with psoriasis immunopathogenesis for assured peptide regions regulator of skin immune response and autoantigen. However, it been a major target proliferation causes high expression in NSCLS that associated poor prognosis and overexpressed the regulation of protein synthesis in a few types of cancer [23].

The characterized of squamous cell carcinoma dependently of site origin through the stratified epithelial expression of KRT5, and KRT17. These studies are required to clarify the mechanisms accomplishment of these biomarkers in the development and progression of lung cancer. Besides, keratin utmost importance for the identification tumor and the selection of treatment plans [16]. The ATP in mitochondrial activates the PI3K-AKT pathway by P2 receptor in some of the cell types, with directly associate the p85 regulation subunit leads to activate the AKT/mTOR pathway [22]. KRT16 expressed as hyper-proliferative keratinocyte-type [16] will act in early inflammatory regulation such as epithelial cancer and chronic inflammatory disorders. The keratinocyte function senses the epidermal barrier of the inflammatory response produces DAMPs as well as attack and attracted the pathogen and immunization cell. It sensitivity affected of inflammatory stimulation the hemidesmosome integrin and keratinocyte. Thus, activation EGF receptor may increase IL-1α in proliferation and inflammatory keratinocytes [24].

The PPBP downregulated gene is a pro-platelet basic protein or CTAP-III derived the platelet growth in chemokine family ligand 7, increases the cancer tissue and blood samples in diagnosis biomarker lung carcinoma [25]. LA-PF4 stimulated by DNA synthesis, mitosis, glycolysis, intracellular cAMP accumulation, prostaglandin E2 secretion, and hyaluronic acid synthesis, and sulfated glycosaminoglycan. Hence, human synovial cells activate the formation and secretion of plasminogen. CXCR1 and CXCR2 ligands have a neutrophil chemoattractant and activator of NAP-2.In vitro, TC-1 and TC-2 are antibacterial proteins released through platelet activated alpha-granules. Activation of neutrophils

or endogenous ligands induced by chemokine and released the activation platelet in large amounts for the response the vascular injury [26]. The hematoregulatory chemokine significantly suppresses hematopoietic progenitor cell proliferation [27] with GRO-beta in the invasion of hepatocellular carcinoma cells while GRO-beta showed a high for enhanced hematopoietic activity.

CXCL1 and CXCL2 bind with target protein of G protein-coupled membrane receptor CXCR2 for immunization in infection site (neutrophils, macrophages, and epithelium cell). It associated the chemotaxis of neutrophils and helps endothelial cells for the growth of the tumor and angiogenesis. Both easy found when attached to CXCL8 that famously as interleukin-8 receptor beta [28]. The receptor chemokine is the combination of the 7-transmembrane domains and GPCRs.

CCR9 is the receptor for chemokine SCYA25/TECK and promoted the cancer cell by inhibited the carcinoma cell. The only ligand for CCR9 is CCL25 that promoted the thymus of chemokine and epithelial cell to the small intestine. When raising the level of intracellular calcium ions, it transduces a signal which same with G-protein coupled receptor 1 family of CCR7. Expression the chemokine of CCR9 affected intrusion the tumor when inhibiting the cytotoxic tumor-infiltrating lymphocytes. And also role for an advance of T lymphocyte and distribution of intestinal surface and immature T lymphocytes.

It works as mediated for the anti-apoptotic progression of lung cancer [29]. The chemokine of CCR9 receptors helps to activate integrin and cytoskeletal reorganization whereas affected cell growth, tumorigenesis, and angiogenesis. The activation of AKT phosphorylation and PI3K may high simultaneously the apoptotic the protein. In [30] stated the silenced of CCR9 in mice displayed it reduced the burden of tumor when it induces the PI3K inhibitor through this silencing. However, the interaction of CCR9 may induce the tumor, especially in NSCLC caused by activation of two inhibitors of PI3K and protein Kinase B.

The activation C5 convertase is spontaneous initiation assembled the last complement components (C5-C9) to attack membrane complex. The C3 and PZP-like contained the domain of alpha-2-macroglobulin in converted C3 molecules to C3a and C3b. The C5 involved in the cascade complement system launches the acute inflammatory responses through activation the C5a and C5b. C5b outbreak the membrane attack complex once with C6-C9. However, the target cell of C5 factor inflammatory and generate the DAMPs and produced T cell of proliferation and survival whereas T cell inhibited the antigen in cell death [31]. Hence, C5 very powerful for anaphylatoxin and inflammation of neutrophils that binding with C5aR lowest apoptosis reaction [32].

The prognostic values assessed the individual genes by used KM plotter of lung cancer patients database shown chemokine ligand types are involved in the corresponding receptor for cancer progression [33]. However, not only this type chemokine, the CCR7 also displayed as significance worse OS. These chemokines are famously known as angiogenesis, regulated immunity, and cell trafficking when binding to corresponding receptors. This organ-specific chemokine receptor's aberrant functional cell surface expression associated with metastasis at this site and receptor function regulation appears to be a critical step in the metastatic

process. Thus, cytokine network lung cancer can be as tumor and anti-cancer immunity microenvironment can affect the chronic inflammation which promotes metastasis and tumor growth.

CONCLUSION

The top250 DEGs in GSE84797, GSE28827, and GSE115456 of gene expression profiling lung cancer involved 365 upregulated and 288 downregulated. These integrated from the bioinformatics method. The PPI network by Cytoscape and STRING integrated the top 5 hub genes in lung cancer which through Cytohubba. The top 5 key genes downregulated were KRT5, KRT13, KRT17, KRTI6, and KRTI5, and upregulated were C5, PPBP, CXCL2, CCR9, and CCR7. The worsen OS of KRT16 and CCR9 role for pathological prognostic with the inflammatory process for tumor growth of lung cancer .C5 mRNA high expression correlated to the good OS in all lung cancer patients. These are helpful for better understanding the biomarker hub genes in molecular biology lung cancer with the accurate tools for expected prognosis. It was also helpful for depth understanding of the mechanism of lung cancer for finding the risk assessment, drug development, and treatment finding.

ACKNOWLEDGEMENT

The researcher would like to take this opportunity to thank the Management and Science University for providing the necessary facilities and encouragement to carry out this work that gives pleasure to acknowledge this research.

REFERENCES

1. Moorthy, B., Chu, C., & Carlin, D. J. (2015). *Polycyclic Aromatic Hydrocarbons□: From Metabolism to Lung Cancer*. *145*(1), 5–15. https://doi.org/10.1093/toxsci/kfv040
2. Herbst, R. S., Morgensztern, D., & Boshoff, C. (2018). *Review The biology and management of non-small cell lung cancer*. https://doi.org/10.1038/nature25183
3. Sharma, P., Mehta, M., Singh, D., Kaur, S., Gupta, G., Singh, H., … Satija, S. (2019). Chemico-Biological Interactions Emerging trends in the novel drug delivery approaches for the treatment of lung cancer. *Chemico-Biological Interactions*, *309*(June), 108720. https://doi.org/10.1016/j.cbi.2019.06.033
4. Lemjabbar-alaoui, H., Hassan, O., Yang, Y., & Buchanan, P. (2015). AC. *BBA - Reviews on Cancer*. https://doi.org/10.1016/j.bbcan.2015.08.002
5. Sun, C., Yuan, Q., Wu, D., Meng, X., & Wang, B. (2017). Identification of core genes and outcome in gastric cancer using bioinformatics analysis. *Oncotarget*, *8*(41), 70271–70280. https://doi.org/10.18632/oncotarget.20082
6. Barrett, T., Wilhite, S. E., Ledoux, P., Evangelista, C., Kim, I. F., Tomashevsky, M., … Soboleva, A. (2013). *NCBI GEO□: archive for functional genomics data sets — update*. *41*(November 2012), 991–995. https://doi.org/10.1093/nar/gks1193
7. Huang, Y., Tao, Y., Li, X., Chang, S., Jiang, B., Li, F., & Wang, Z. M. (2017). Bioinformatics analysis of key genes and latent pathway interactions based on the anaplastic thyroid carcinoma gene expression profile. *Oncology Letters*, *13*(1), 167–176. https://doi.org/10.3892/ol.2016.5447
8. Huang, D. W., Sherman, B. T., Tan, Q., Kir, J., Liu, D., Bryant, D., … Lempicki, R. A. (2007). DAVID Bioinformatics Resources: Expanded annotation database and novel

algorithms to better extract biology from large gene lists. *Nucleic Acids Research*, *35*(SUPPL.2), 169–175. https://doi.org/10.1093/nar/gkm415
9. Szklarczyk, D., Gable, A. L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., … Von Mering, C. (2019). STRING v11: Protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Research*, *47*(D1), D607–D613. https://doi.org/10.1093/nar/gky1131
10. Chin, C. H., Chen, S. H., Wu, H. H., Ho, C. W., Ko, M. T., & Lin, C. Y. (2014). cytoHubba: Identifying hub objects and sub-networks from complex interactome. *BMC Systems Biology*, *8*(4), 1–7. https://doi.org/10.1186/1752-0509-8-S4-S11
11. Gyorffy, B., Surowiak, P., Budczies, J., & Lánczky, A. (2013). Online survival analysis software to assess the prognostic value of biomarkers using transcriptomic data in non-small-cell lung cancer. *PLoS ONE*, *8*(12). https://doi.org/10.1371/journal.pone.0082241
12. Chen, L., Huang, C., Yang, X., Zhang, Q., & Chen, F. (2018). Prognostic roles of mRNA expression of peroxiredoxins in lung cancer. *OncoTargets and Therapy*, *11*, 8381–8388. https://doi.org/10.2147/OTT.S181314
13. Tang, Z., Kang, B., Li, C., Chen, T., & Zhang, Z. (2019). GEPIA2: an enhanced web server for large-scale expression profiling and interactive analysis. *Nucleic Acids Research*, *47*(W1), W556–W560. https://doi.org/10.1093/nar/gkz430
14. Zhang, M. L., Sun, W. H., Wu, H. Q., Liu, Z. D., & Wang, P. (2020). Knockdown of microRNA-103a-3p inhibits the malignancy of thyroid cancer cells through Hippo signaling pathway by upregulating LATS1. *Neoplasma*, *67*(6), 1266–1278. https://doi.org/10.4149/neo_2020_191224N1331
15. Ling, B. O., Liao, X., Huang, Y., & Liang, L. (2020). Identification of prognostic markers of lung cancer through bioinformatics analysis and in vitro experiments. 193–205. https://doi.org/10.3892/ijo.2019.4926
16. Bergendal, E. (2008). 基因的改变NIH Public Access. *Bone*, *23*(1), 1–7. https://doi.org/10.1038/onc.2010.456.Keratins
17. Eckstein, M., Wirtz, R. M., Gross-Weege, M., Breyer, J., Otto, W., Stoehr, R., … Erben, P. (2018). Mrna-expression of krt5 and krt20 defines distinct prognostic subgroups of muscle-invasive urothelial bladder cancer correlating with histological variants. *International Journal of Molecular Sciences*, *19*(11). https://doi.org/10.3390/ijms19113396
18. Li, Q., Yin, L., Jones, L. W., Chu, G. C. Y., Wu, J. B. Y., Huang, J. M., … Zhau, H. E. (2016). Keratin 13 expression reprograms bone and brain metastases of human prostate cancer cells. *Oncotarget*, *7*(51), 84645–84657.
19. Bose, A., Teh, M. T., Mackenzie, I. C., & Waseem, A. (2013). Keratin K15 as a biomarker of epidermal stem cells. *International Journal of Molecular Sciences*, *14*(10), 19385–19398. https://doi.org/10.3390/ijms141019385
20. Lastwika, K. J., Wilson, W., Li, Q. K., Norris, J., Xu, H., Ghazarian, S. R., … Dennis, P. A. (2016). Control of PD-L1 expression by oncogenic activation of the AKT-mTOR pathway in non-small cell lung cancer. *Cancer Research*, *76*(2), 227–238. https://doi.org/10.1158/0008-5472.CAN-14-3362
21. Liu, Z., Yu, S., Ye, S., Shen, Z., Gao, L., Han, Z., … Kang, M. (2020). Keratin 17 activates AKT signalling and induces epithelial-mesenchymal transition in oesophageal squamous cell carcinoma. *Journal of Proteomics*, *211*, 103557. https://doi.org/10.1016/j.jprot.2019.103557
22. Yu, T., Zhao, Y., Hu, Z., Li, J., Chu, D., Zhang, J., … Yao, M. (2017). Meta Lnc9 facilitates lung cancer metastasis via a PGK1-activated AKT/mTOR pathway. *Cancer Research*, *77*(21), 5782–5794. https://doi.org/10.1158/0008-5472.CAN-17-0671
23. Wang, Z., Yang, M. Q., Lei, L., Fei, L. R., Zheng, Y. W., Huang, W. J., … Xu, H. T. (2019). Overexpression of KRT17 promotes proliferation and invasion of non-small cell lung cancer and indicates poor prognosis. *Cancer Management and Research*,

11, 7485–7497. https://doi.org/10.2147/CMAR.S218926
24. Lessard, J. C., Piña-paz, S., Rotty, J. D., Hickerson, R. P., Kaspar, R. L., & Balmain, A. (2013). *Keratin 16 regulates innate immunity in response to epidermal barrier breach*. *110*(48), 19537–19542. https://doi.org/10.1073/pnas.1309576110
25. Shi, K., Li, N., Yang, M., & Li, W. (2019). *Identification of Key Genes and Pathways in Female Lung Cancer Patients Who Never Smoked by a Bioinformatics Analysis*. *10*. https://doi.org/10.7150/jca.26908
26. Maneerat, Y., Prasongsukarn, K., Benjathummarak, S., & Dechkhajorn, W. (2017). *PPBP and DEFA1 / DEFA3 genes in hyperlipidaemia as feasible synergistic inflammatory biomarkers for coronary heart disease*. 1–12. https://doi.org/10.1186/s12944-017-0471-
27. Lv, J., & Li, L. (2019). *Hub Genes and Key Pathway Identification in Colorectal Cancer Based on Bioinformatic Analysis*. *2019*.
28. Chen, M. C., Hsu, H. H., Lin, Y. M., & Viswanadha, V. P. (2018). *CXCL2 / CXCR2 axis induces cancer stem cell characteristics in CPT - 11 - resistant LoVo colon cancer cells via G α i - 2 and G α q / 11*. (June), 1–13. https://doi.org/10.1002/jcp.27891
29. Tu, Z., Xiao, R., Xiong, J., Tembo, K. M., Deng, X., Xiong, M., … Wang, M. (2016). CCR9 in cancer□: oncogenic role and therapeutic targeting. *Journal of Hematology & Oncology*, *3*, 1–9. https://doi.org/10.1186/s13045-016-0236-7
30. Xu, B., Deng, C., Wu, X., Ji, T., Zhao, L., Han, Y., … Yang, Y. (2020). CCR9 and CCL25: A review of their roles in tumor promotion. *Journal of Cellular Physiology*, *235*(12), 9121–9132. https://doi.org/10.1002/jcp.29782
31. Savola, S., Klami, A., Myllykangas, S., Manara, C., Scotlandi, K., Picci, P., … Vakkila, J. (2011). *High Expression of Complement Component 5 (C5) at Tumor Site Associates with Superior Survival in Ewing ' s Sarcoma Family of Tumour Patients*. *2011*. https://doi.org/10.5402/2011/168712
32. Afshar-kharghan, V. (2017). *The role of the complement system in cancer*. *127*(3).
33. Zheng, Z., Cai, Y., Chen, H., Chen, Z., Zhu, D., Zhong, Q., & Xie, W. (2019). CXCL13/CXCR5 axis predicts poor prognosis and promotes progression through PI3K/AKT/mTOR pathway in clear cell renal cell carcinoma. *Frontiers in Oncology*, *9*(JAN), 1–10. https://doi.org/10.3389/fonc.2018.00682

TABLES

Table 1: Patients' information in GSE84797, GSE28827, and GSE115456 derived from the GEO database. 14

Sample	Total Sample	Type
GSE84797	28	Primary lung cancer
GSE28827	26	Non-small cell lung carcinoma
GSE115456	30	Non-small cell lung carcinoma

Table 2: Gene Ontology (GO) terms such as biological process, molecular functions, and cellular component of DEGs by using DAVID v6.8 online tools. 23

Expression	Category	Term	count	%	p-value	FDR
Down-regulated	GOTERM_BP_DIRECT	Immune response	43	15.3	5.10E-22	8.70E-19
	GOTERM_BP_DIRECT	Inflammatory response	30	10.7	1.90E-12	1.60E-09
	GOTERM_BP_DIRECT	Regulation of complement activation	10	3.6	5.50E-10	1.90E-07
	GOTERM_BP_DIRECT	Chemotaxis	14	5	6.40E-08	1.60E-05
	GOTERM_BP_DIRECT	Chemokine-mediated signaling pathway	10	3.6	1.60E-06	2.50E-04
	GOTERM_MF_DIRECT	Protein binding	156	55.5	9.70E-03	1.90E-01
	GOTERM_MF_DIRECT	Chemokine activity	11	3.9	3.20E-09	1.30E-06
	GOTERM_MF_DIRECT	Receptor binding	15	5.3	1.20E-03	3.70E-02

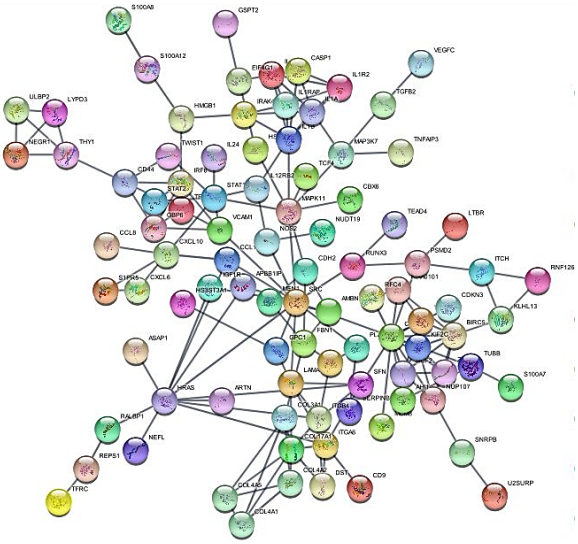
	GOTERM_CC_DIRECT	Plasma membrane	116	41.3	1.10E-13	2.80E-11
	GOTERM_CC_DIRECT	Extracellular space	52	18.5	3.70E-10	3.10E-08
	GOTERM_CC_DIRECT	Cell surface	30	10.7	2.30E-09	1.40E-07
	GOTERM_CC_DIRECT	Extracellular region	50	17.8	8.20E-07	2.90E-05
	GOTERM_CC_DIRECT	Extracellular exosome	69	24.6	1.70E-05	4.20E-04
Up-regulated	GOTERM_BP_DIRECT	Response to lipopolysacch-aride	13	3.6	7.80E-05	2.10E-02
	GOTERM_BP_DIRECT	Epidermis development	8	2.2	1.20E-03	1.50E-01
	GOTERM_BP_DIRECT	Intermediate filament cytoskeleton organization	4	1.1	2.70E-03	2.30E-01
	GOTERM_BP_DIRECT	Cytoskeleton organization	10	2.8	3.90E-03	2.70E-01
	GOTERM_BP_DIRECT	Innate immune response	18	5	4.00E-03	2.70E-01
	GOTERM_MF_DIRECT	Structural constituent of cytoskeleton	10	2.8	4.00E-03	2.70E-01
	GOTERM_MF_DIRECT	Scaffold protein binding	7	2	2.90E-04	2.80E-02
	GOTERM_MF_DIRECT	Protein binding	11	3.1	2.10E-03	1.00E-01
	GOTERM_MF_DIRECT	Structural molecule activity	11	3.1	2.10E-02	7.40E-01
	GOTERM_CC_DIRECT	extracellular exosome	79	22.1	6.70E-05	1.10E-02
	GOTERM_CC_DIRECT	intermediate filament	9	2.5	1.10E-03	6.90E-02

Table 3: KEGG pathway analysis of differentially expressed genes associated with lung cancer by using DAVID v6.8 online tools 19

Expression	Term	count	%	p-value	FDR
Down-regulated	Cytokine-cytokine receptor interaction	29	10.3	2.00E-12	3.10E-10
	Herpes simplex infection	20	7.1	5.40E-08	1.20E-06
	Intestinal immune network for IgA	11	3.9	1.10E-07	2.10E-06

	production				
	Complement and coagulation cascades	11	3.9	4.60E-06	4.40E-05
	Chemokine signaling pathway	17	6	7.90E-06	7.10E-05
	Pertussis	10	3.6	6.50E-05	4.50E-04
Upregulated	Focal adhesion	13	3.6	2.50E-03	6.70E-02
	Leishmaniasis	7	2	5.40E-03	1.00E-01
	Small cell lung cancer	6	1.7	4.40E-02	4.10E-01

FIGURES



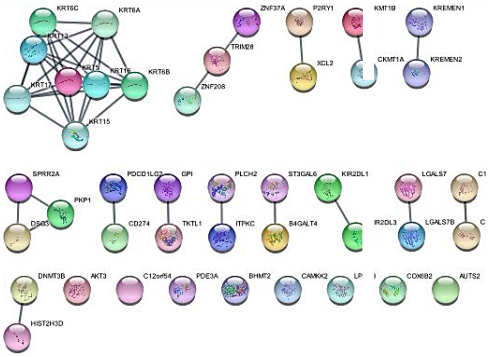


Figure 1: The protein-protein network of upregulated lung cancer with 338 nodes and 212 edges. Diagram obtained from Cytoscape v3.8.1 database.

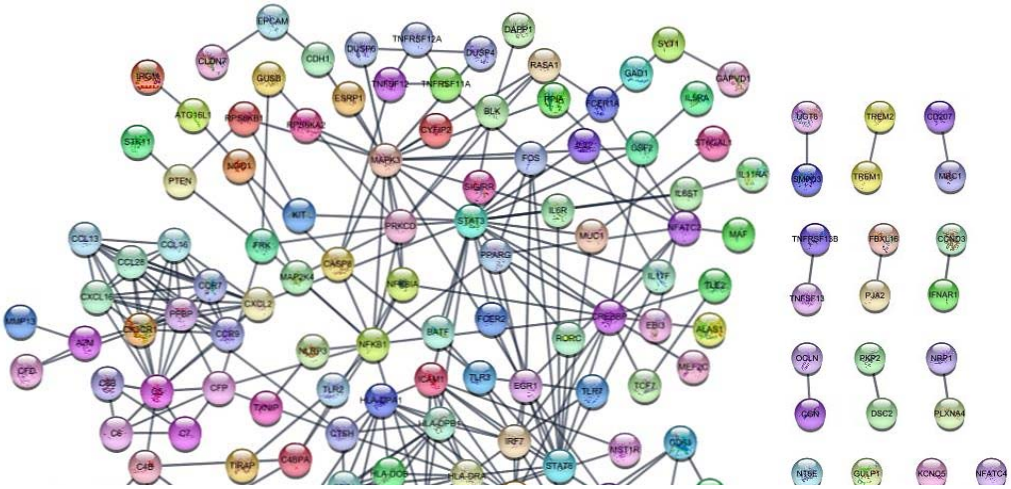


Figure 2: The protein-protein network of downregulated lung cancer with 277 nodes and 276 edges. Diagram obtained from Cytoscape v3.8.1 database.

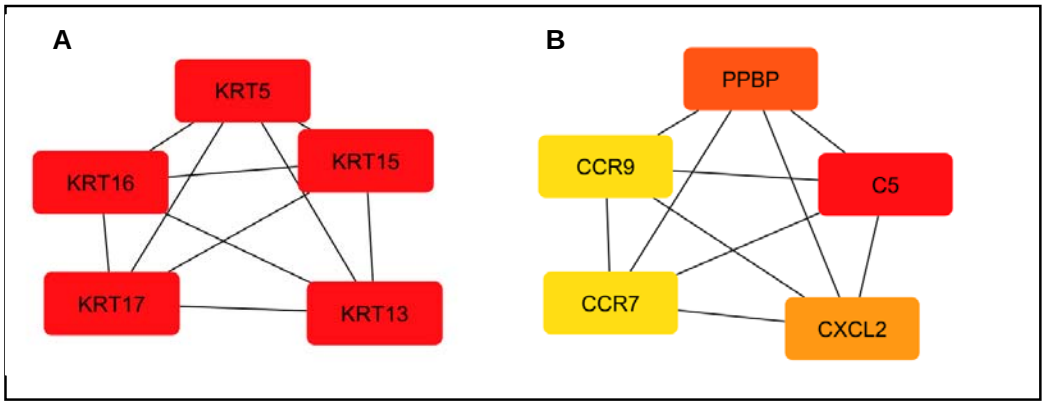


Figure 3: Top 5 hub genes (A) upregulated (B) downregulated of lung cancer. Diagram obtained from cytoHubba plugin Cytoscape v3.8.1

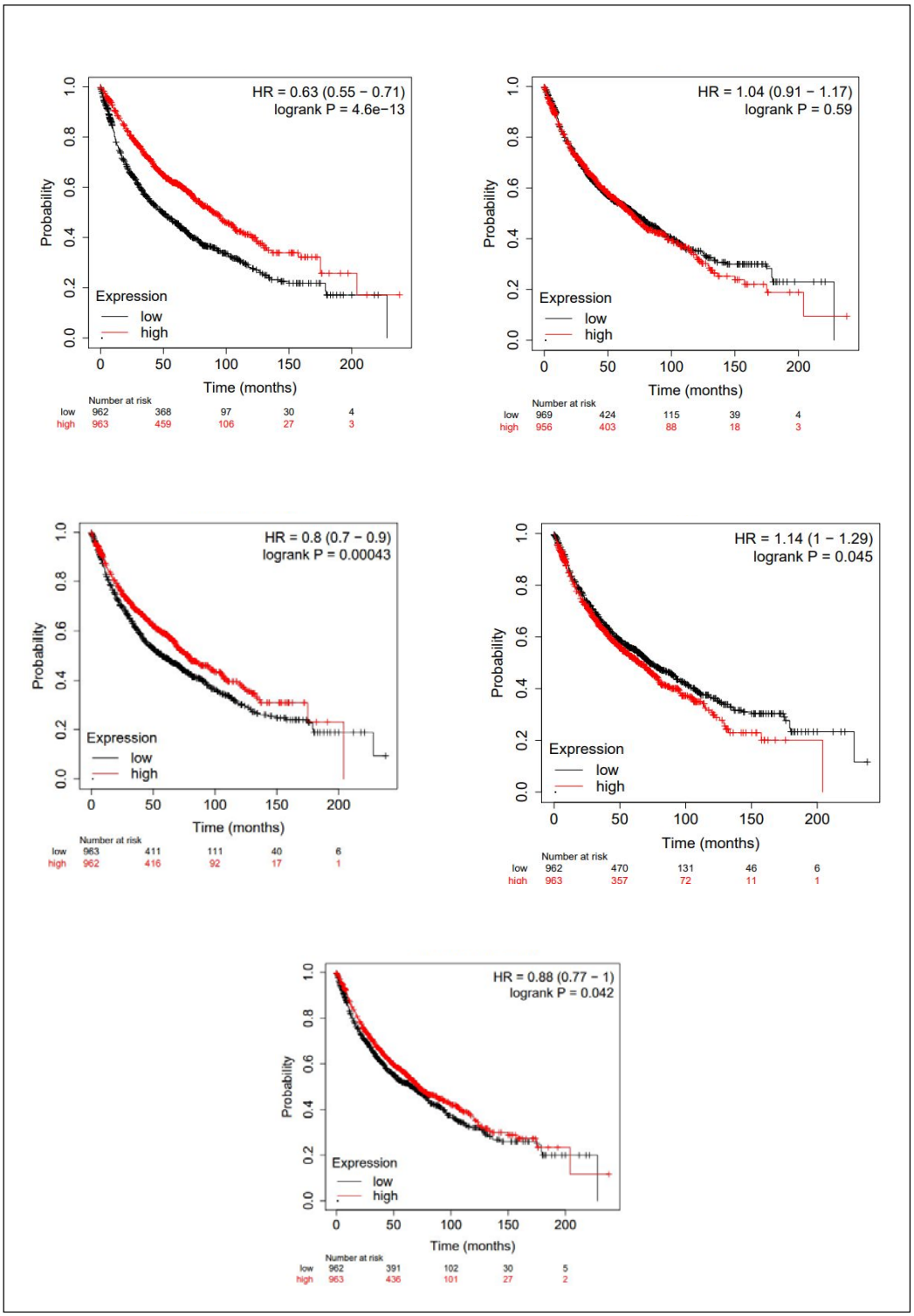


Figure 4: Prognostic value of five hub genes (A) C5, (B) PPBP, (C) CCR7, (D) CCR9, (E) CXCL2 in downregulated of lung cancer. The desired Affymetrix IDs are valid 205500_at (C5), 214146_s_at (PPBP), 206337_at (CCR7), 207445_s_at (CCR9),209974_x_at (CXCL2). HR: hazard ratio, CI: confidence interval. Diagram obtained from Kaplan-Meier Plotter online server

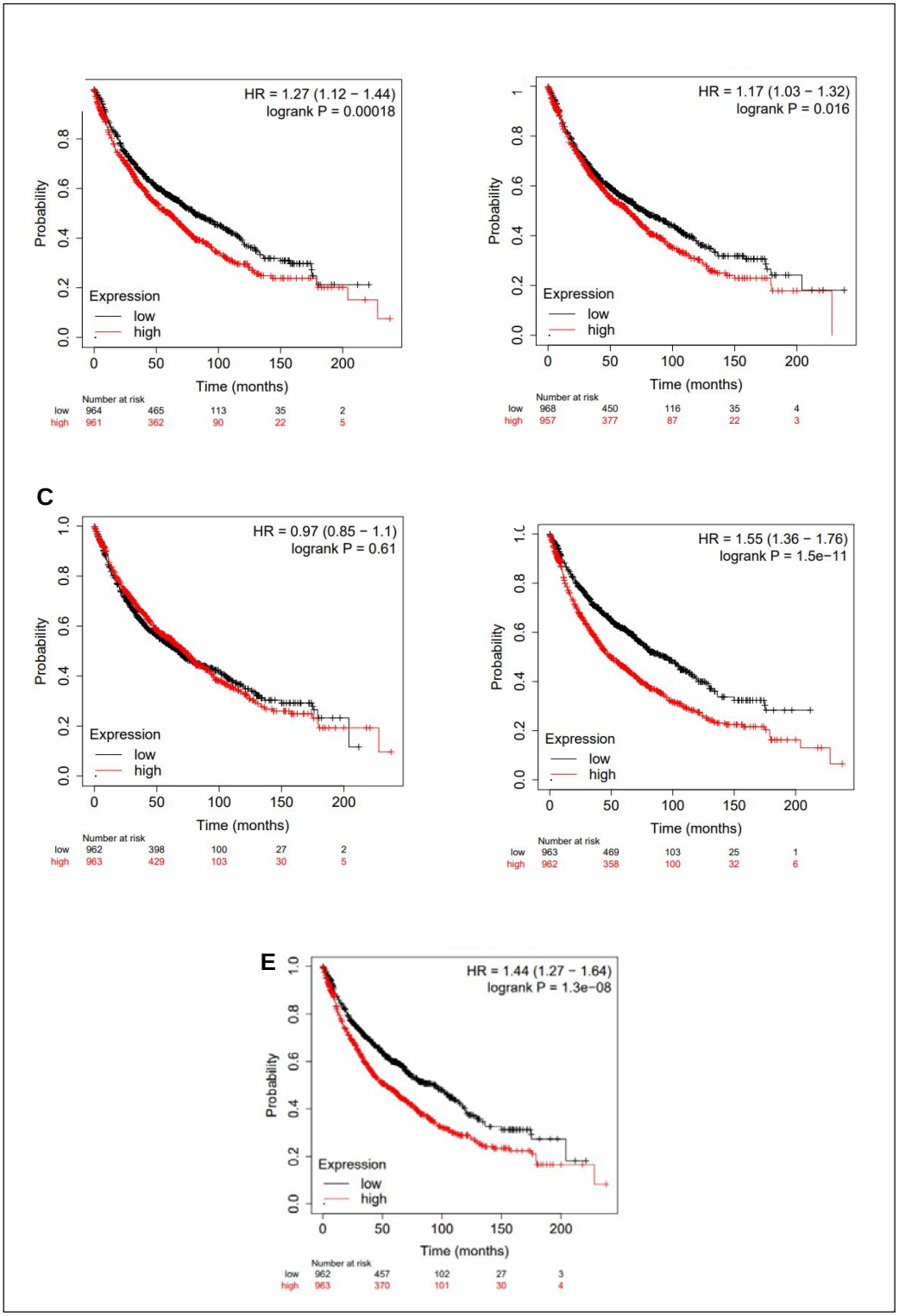


Figure 5: Prognostic value of five genes (A) KRT5, (B) KRT13, (C) KRT15, (D) KRT16, (E) KRT17 in upregulated of lung cancer. The desired Affymetrix IDs are valid 201820_at (KRT5), 207935_s_at (KRT13), 204734_at (KRT15), 209800_at (KRT16), 205157_s_at (KRT17). HR: hazard ratio, CI: confidence interval. Diagram obtained from KM plotter online servers

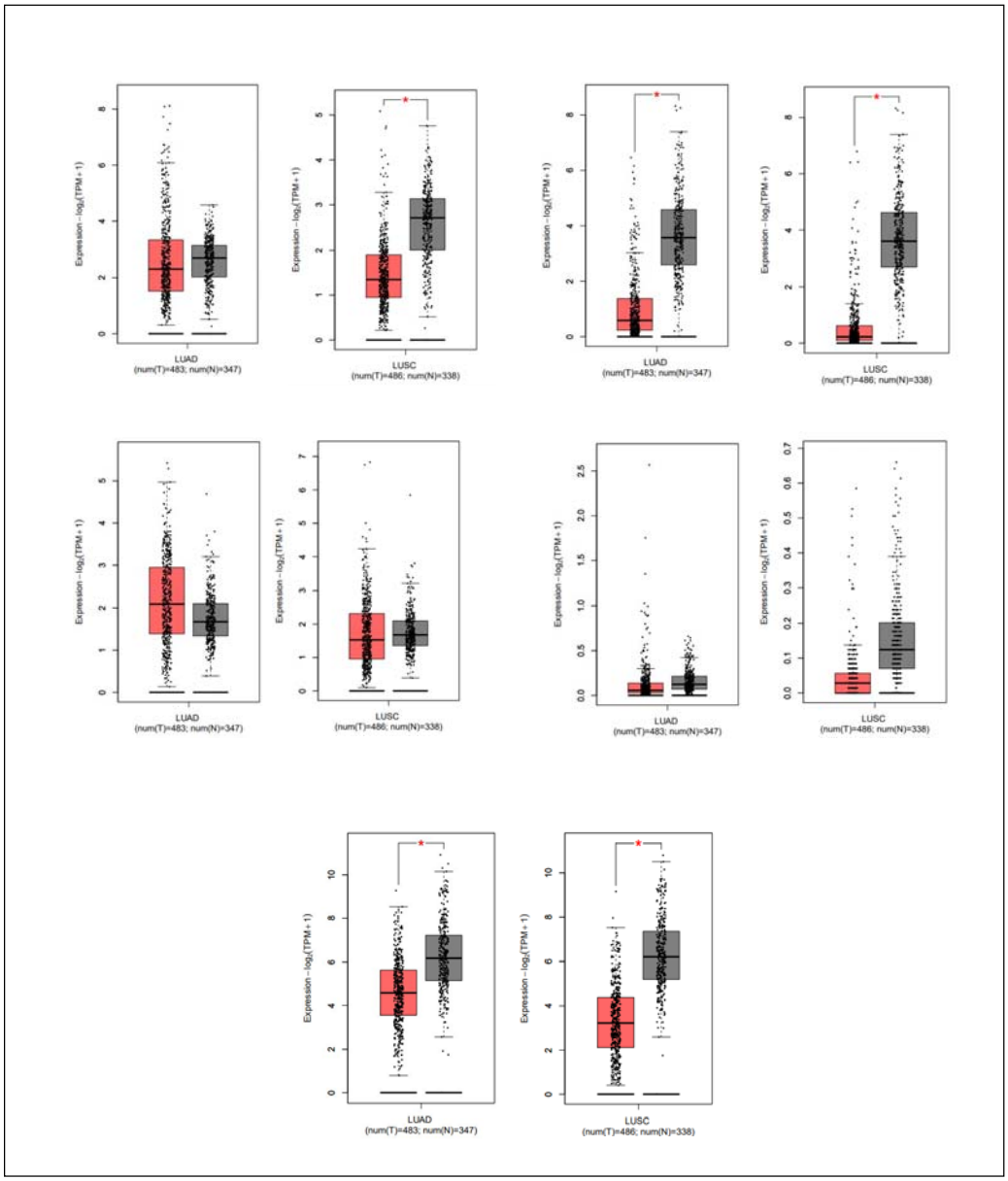


Figure 6: Expression level of five hub genes (A)C5, (B)PPBP, (C) CCR7, (D) CCR9, (E)CXCL2 in downregulated of lung cancer. Diagram obtained from GEPIA2 online server.

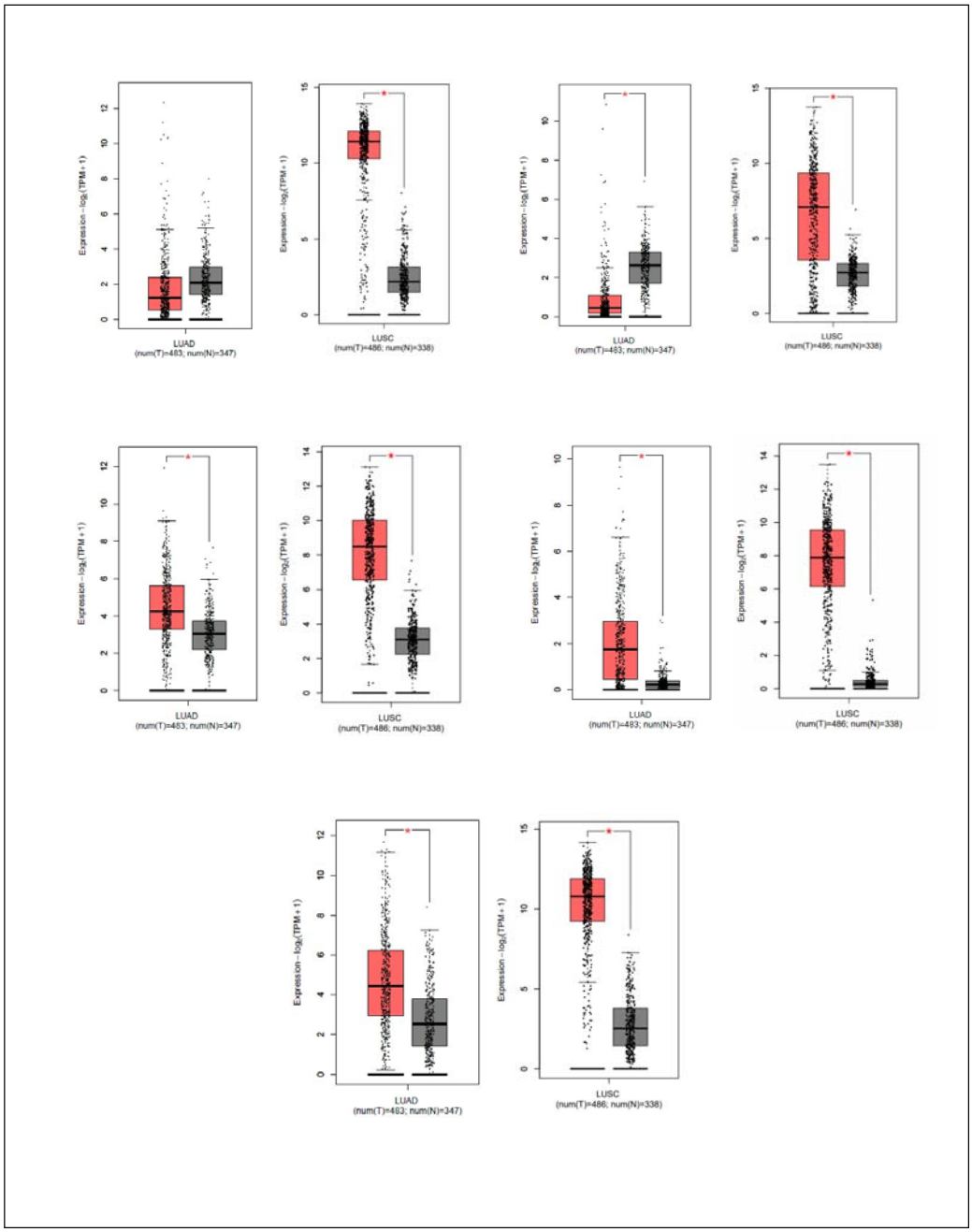


Figure 7: Expression level of five genes (A) KRT5, (B) KRT13, (C) KRT15, (D) KRT16,(E)KRT17 in upregulated of lung cancer. Diagram obtained from GEPIA2 online server.