Genes co-related with poor prognosis of patients with lung cancer via bioinformatical approaches

Dai Shi
Shandong University https://orcid.org/0000-0003-3903-5713
Yeming Han
Qilu Hospital
Xin Wang
Shandong University
Guihua Hou (ghhou@sdu.edu.cn)
Shandong University https://orcid.org/0000-0001-7211-0709

Research

Keywords: lung cancer, bioinformatical analysis, GEO

Posted Date: March 19th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-17941/v1

Abstract

Lung cancer is one of the most common malignant tumors with high mortality worldwide. Recently, researchers reported that molecular markers on lung cancer could be used as diagnostic and prognostic targets. However, these molecules were not ideal in specificity and high selectivity. Therefore, exploring more reliable biomarkers to improve the prognosis and clarify the underlying mechanism is urgently needed both for clinic and basic research. This study aimed to identify significant genes with poor prediction for lung cancer and their underlying mechanisms. Firstly, we used gene expression datasets available from GEO (Gene Expression Omnibus) database. There were 109 lung cancer samples and 27 normal samples in the selected datasets. First, DEGs (Different Expressed Gene set) of lung cancer and normal lung samples were screen out with GEO2R tool, and we displayed them by Venn diagram software and Heatmap. Secondly, we used DAVID (Database for Annotation, Visualization and Integrated Discovery) to analyze KEGG (Kyoto Encyclopedia of Gene and Genome) pathway and GO (Gene Ontology). Third, PPI (Protein-Protein Interaction) of these DEGs was conducted by Cytoscape with STRING (Search Tool for the Retrieval of Interacting Genes). Our results showed that the expression trends of 21 up-regulated genes and 116 down-regulated were similar in selected three datasets. Analyzed by MCODE (Molecular Complex Detection) plug-in, 11 up-regulated and 16 down-regulated genes were selected. To further verify gene expression differences, GEPIA (Gene Expression Profiling Interactive Analysis) was implemented and we found 26 of 27 genes were found differently expressed in lung cancer compared with normal lung tissues. Furthermore, Kaplan–Meier analysis was used and we found 23 of 26 genes for overall survival indicated much less survival time. At last, three genes, CDH5, CLDN5, PECAM1, were found to be significantly decreased in lung cancer tissue proved through re-analysis of DAVID, which mainly co-related with leukocyte trans-endothelial migration. In conclusion, three significant down-regulated deferentially expressed genes with poor prognosis on lung cancer were identified basing on integrated bioinformatical methods. These down-expressed genes may be as a potential prognosis targets for patients with lung cancer.

Introduction

Lung cancer is the most common malignant tumor with high mortality both in male and female around the world [1–3]. Due to the occultation of lung cancer, most patients have been in the advanced stage at the time of diagnosis [4]. Regardless of different subtypes, the overall survival rate of lung cancer patients is still disappointing; less than 7% of patients survived 10 years following diagnosis across all stages of lung cancer [5]. Recently, researchers reported that molecular markers on lung cancer cells could be used as diagnostic target, such as 14C5 and α3β1 [6–7]. However, these molecules were not ideal in specific and high selective properties. Therefore, exploring more reliable biomarkers to improve the prognosis and clarifying the underlying mechanism is urgently needed both for clinic and basic research.

Recently, researchers reported that molecular markers on lung cancer cells could be used as diagnostic and prognostic targets. However, these molecules were not ideal and suitable for clinic application.

Gene chip assay was already applied in gene expression [8]. However, only changes in one or more amount of genes could be analyzed simultaneously with this method. Recently, bioinformatical methods were reported which could be used among multiple genes to further investigate the underlying mechanisms, signal pathway and the interaction [9].

In this study, firstly, we chose GSE33532, GSE43346 and GSE 118370 from GEO database. Secondly, we applied for GEO2R online tool and displayed DEGs by Venn diagram software and Heatmap. Thirdly, the DAVID was used to analyze these DEGs, including GO analysis (molecular function (MF), cellular component (CC), biological process (BP)) and KEGG pathways. The fourth, to find out core genes, we established PPI network and then implemented Cytotype MCODE for additional analysis of the DEGs. Then these core genes were imported into the Kaplan Meier plotter online database and GEPIA for the significant survival information and further verifying the different expression($P < 0.05$). Taken all data above, 23 DEGs were screened out. Then, we re-analyzed these 23 DEGs for KEGG pathway enrichment. At last, three genes (CDH5, CLDN5, PECAM1) were obtained and significantly enriched in the Leukocyte trans-endothelial migration.

Methods

Microarray data information

GSE33532, GSE43346 and GSE 118370, three gene expression profiles about lung cancer samples and normal lung samples, were obtained from NCBI-GEO [10]. Microarray data of GSE33532, GSE43346 and GSE 118370 was all on account of GPL570 Platforms, including 80 lung cancer samples and 20 normal lung samples, 23 lung cancer samples and 1 normal lung samples, and 6 lung cancer samples and 6 normal lung samples, respectively.

Screen out DEGs

DEGs between lung cancer and normal lung samples of three datasets were screened out by GEO2R online tool [11] with $|logFC| > 2$ and p value < 0.05. That DEGs screened out by $logFC < -2$ were considered as down-regulated genes. On the contrary, the DEGs which were screened out by $logFC > 2$ were considered as up-regulated genes.

Draw Venn diagram and heatmap

Venn diagram and heatmap were respectively drawn by the tool of Venn website (http://bioinformatics.psb.ugent.be/webtools/Venn/) and Excel software.

Gene ontology analysis and KEGG analysis by DAVID

Gene ontology analysis was an approach, which was commonly used in defining genes, relative RNA and relative protein to identify unique biological properties of high-throughput transcriptome or genome data [12]. KEGG is an encyclopedia of genes and genomes [13]. DAVID, a functional annotation tool, was

conducted to identify function for tremendous genes, RNA and proteins [14]. Here, our study used DAVID to verify the DEGs' enrichment and their underlying pathways.

PPI networks and core genes

We used STRING, a kind of functional protein association networks, for evaluating PPI information [15]. Then, Cytoscape software [16] was applied to find out the potential correlation between these DEGs. Furthermore, the MCODE app of Cytoscape software was used to screen out the core genes of the PPI network.

Verification of gene expression differences and survival analysis of core genes

To further verify gene expression differences, we used the GEPIA website (http://gepia.cancer-pku.cn/) for analyzing RNA sequencing expression from the GTEx projects and TCGA, which included thousands of samples [17]. Kaplan Meier-plotter (http://www.kmplot.com/lung/) [18], a survival analysis tool, widely used in assessing the survival information of a large number of differently expressed genes based on EGA, GEO and TCGA database. The logrank *p* value (<0.05) and hazard ratio with 95% confidence intervals were shown on the upper right corner of plots.

Results

Identification of DEGs in lung cancer samples compared to normal lung

There were totally 109 lung cancer samples and 27 normal lung samples in our study. Analyzing by GEO2R online tool, we screened out 611, 2880 and 1071 DEGs from GSE 33532, GSE 43346 and GSE 118370, respectively. Then, we used Venn diagram software to screen out the corporate DEGs in the three datasets. Results showed there were totally 137 corporate DEGs, including 21 up-regulated genes (logFC>2) and 116 down-regulated genes (logFC<-2) in the lung cancer samples compared to normal lung samples (Table 1 & Fig 1).

GO analysis and KEGG pathway analysis of DEGs in lung cancers

Totally, 137 DEGs were analyzed by DAVID 6.7. The relative results of GO analysis were shown in Table 2.

In BP, up-regulated DEGs were mainly enriched in different stages of cell cycle, while down-regulated DEGs were particularly mainly enriched in vasculature development, blood vessel morphogenesis and so on.

In CC, up-regulated DEGs were mainly enriched in microtubule, cytoskeleton, non-membrane-bounded organelle and so on. Down-regulated DEGs were mainly enriched in plasma membrane part and so on.

In MF, down-regulated DEGs were mainly enriched in transforming growth factor beta bin and actin filament binding.

Results about KEGG analysis were shown in Table 3. From the Table 3, We found that up-regulated DEGs were mainly enriched in p53 signaling pathway, while down-regulated DEGs were mainly enriched in vascular smooth muscle contraction, cell adhesion molecules, dilated cardiomyopathy and so on.

PPI networks and core genes analysis

The expression of total 137 DEGs was shown in heatmap, which was shown the relative expression level of DEGs(Fig. 2A)[19-20]. Then the 137 DEGs were verified by the DEGs PPI network tool of STRING, which included 243 edges and 104 nodes. And there were 87 down-regulated and 17 up-regulated genes after analyzing by PPI network (Fig. 2B). Then we conducted Cytotype MCODE app for a further analysis. the Fig. 2C-D showed that 27 central genes which were consisted of 11 up-regulated genes and 16 down-regulated genes were screen out.

Analysis of core genes by the Kaplan Meier plotter and GEPIA

GEPIA was used to further verify the 27 core genes' expression between the lung cancer samples and normal lung samples. Our results indicated that 26 of 27 genes indicated high or low expressed in lung cancer samples compared to normal lung samples. ($P<0.05$, Table 4 and Fig. 3).Meanwhile, Kaplan-Meier plotter was implemented in identifying 26 core genes survival information. It was found that there were 23 genes of 26 core genes having significantly worse survival information, while 3 genes were no significant (Table 5 & Fig. 4). So, there were 23 meaningful genes left after GEPIA and Kaplan-Meier plotter analysis.

Re-analysis of 23 genes via DAVID

KEGG pathway enrichment was re-analyzed via DAVID to further find the possible pathway of these 23 core genes. Our results showed that three genes (CDH5, CLDN5, PECAM1) markedly enriched in the leukocyte trans-endothelial migration and CAMs (cell adhesion molecules) (Table 6 & Fig. 5A and 5B). Furthermore, we found there was no correlation among CDH5, CLDN5, PECAM1 verified by Pearson assay ($P>0.05$, respectively).

Discussion

In this study, we performed a bioinformatical analysis on the basis of three gene chip datasets, to find more efficient biomaker of lung cancer. 109 lung cancer samples and 27 normal samples were involved in our study. The results revealed total 137 corporately changed DEGs and analyzed via GEO2R and shown by Venn diagram and heatmap [21]. GO analysis and KEGG pathway enrichment analysis using DAVID methods [22] indicated which pathways up-regulated DEGs or down-regulated DEGs were enriched in. 27 center genes were screen out via the PPT network and Cytoscape MCODE app software [23–24]. Furthermore, through GEPIA analysis [25] and Kaplan-Meier plotter analysis [26], we found that 26 genes showed apparently high or low expression in lung cancer samples compared with normal samples by among these 27 core genes and 23 of 26 genes indicated a significantly worse survival. Finally, we re-

analyzed 23 core genes and found that three genes (CDH5, CLDN5, PECAM1) enriched in the leukocyte trans-endothelial migration and cell adhesion molecules (CAMs), which might be considered as new effective targets to play a role on the diagnosis and prognosis for patients with lung cancer.

CDH5 (cadherin 5), encoded a classical cadherin, located on the long arm of chromosome 16, involved in loss of heterozygosity events in breast and prostate cancer. In 2016, Hung found the CDH5 as an angiogenic factor in lung cancer[27]. Furthermore, not only in lung cancer, Mao reported that CDH5 was overexpressed in gliomas, co-related with tumor grades, and was an independent adverse prognostic predictor for patients with glioblastoma multiforme [28]. In addition, CDH5 was reported that it played a role in regulating angiogenesis, human drug-induced liver injury and gastric cancer [29–31].

CLDN5 (claudin 5) was a member of the claudin family and claudins belong to integral membrane proteins and components of tight junction strands. Mutations in this gene have been found in patients with velocardiofacial syndrome. Jia reported that down-regulating CLDN5 increased tumor invasion and potential metastatic abilities[32]. Ma discovered CLDN5 was closely related to brain metastases from lung cancer[33]. In 2019, Jia indicated that high-dose bevacizumab likely increased lung tumor invasion and potential metastatic abilities through down-regulating CLDN5[34].Moreover, CLDN5 showed a close relationship with mental illness, such as depression [35], schizophrenia [36], brain edema following fatal heat stroke [37] and tumor brain metastasis [38].

PECAM1 (platelet and endothelial cell adhesion molecule) has been found on the surface of platelets, monocytes, neutrophils, and some types of T-cells, and existed in a large portion of endothelial cell inter-cellular junctions. The encoded protein was a member of the immunoglobulin superfamily and involved in leukocyte migration, angiogenesis, and integrin activation. Ilhan-Mutlu A and Frezzetti D showed that PECAM1 was found to be related to angiogenesis of the lung cancer [39–40]. Furthermore, PECAM1 was related to vascular channels [41], cerebral malaria [42], recurrent implantation failure [43], acute myeloid leukemia [44] and human glioblastoma multiforme [45].

Our results suggested that bioinformatical analysis on the basis of gene chip datasets could be used to find more efficient biomaker for lung cancer. However, there were still some limitations in this study. Firstly, the samples size was not big, which might result in some results deviations. Secondly, even though numerous studies proved that these three genes were related to various types of cancer, however, very few studies have been reported about CDH5, CLDN5 and PECAM1 in prognostic evaluation of lung cancer based on Pubmed retrieval. Therefore, our finding may provide useful information for future study about these three genes in lung cancer.

Conclusion

Our study based on bioinformatics analysis identified three down-regulated DEGs (CDH5, CLDN5, PECAM1) with poor prognosis of patients with lung cancer. These decreased expressed genes may be as potential prognosis predicting targets and may be very helpful for clarifying the mechanisms of prognosis for patients with lung cancer.

Declarations

ACKNOWLEDGEMENTS

This work was supported by grants from the National Natural Science Foundation of China (81371601, GH Hou) and The Natural Science Foundation of Shandong Province (ZR2019MH019, to GH Hou)

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

Our study was based on public database and did not require ethical approval. Our study did not involve patient consent.

CONFLICT OF INTEREST

The authors confirm that there are no conflicts of interest.

DATA AVAILABILITY STATEMENT

All data are available upon request.

CONSENT FOR PUBLICATION

Not applicable

AUTHOR CONTRIBUTIONS

Guihua Hou designed research; Dai Shi conducted research; Dai Shi and Yeming Han analyzed date; Yeming Han and Xin Wang revised manuscript. All authors wrote the paper and had primary responsibility for final content. All the authors read and approved the final manuscript.

References

1. Mao Y, Yang D, He J, et al. Epidemiology of Lung Cancer. Surgical Oncology Clinics of North America 2016; 25:439-445.
2. Nanavaty P, Alvarez M S, Alberts W M. Lung Cancer Screening: Advantages, Controversies, and Applications. Cancer Control 2014; 21:9-14.
3. Schwartz AG, Cote ML. Epidemiology of Lung Cancer. Adv Exp Med Biol 2016; 893: 21-
4. Yao Q, Zhang A M , Ma H, et al. Novel molecular beacons to monitor microRNAs in non-small-cell lung cancer . Molecular and Cellular Probes 2012; 26:182-187.
5. Crino L.Weder W.,Van Meerbeeck J,et al. Early stage and locally advanced (non-metastatic) non-small-cell lung cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. Annals of Oncology 2010; 21:v103-v115.

6. Burvenich, I. In vitro and In vivo Targeting Properties of Iodine-123- or Iodine-131-Labeled Monoclonal Antibody 14C5 in a Non-Small Cell Lung Cancer and Colon Carcinoma Model. Clinical Cancer Research 2005; 11:7288-7296.
7. Chen Z, Gao H, Li M, et al. Targeted radionuclide therapy for lung cancer with iodine-131-labeled peptide in a nude-mouse model. Anti-cancer drugs 2017; 28:480.
8. Wang XW, Gao HJ, Fang DC. Advances in gene chip technique in Barrett's metaplasia and adenocarcinoma. J Dig Dis 2008; 9: 68–71.
9. Zhang Y, Wang D C, Shi L, et al. Genome analyses identify the genetic modification of lung cancer subtypes. Seminars in Cancer Biology 2017; 42:20-30.
10. Clough E, Barrett T. The Gene Expression Omnibus Database. Methods in Molecular Biology 2016; 1418:93.
11. Davis S, Meltzer P S. GEOquery: A bridge between the Gene Expression Omnibus (GEO) and BioConductor. Bioinformatics 2007; 23: 1846-1847.
12. Ashburner M, Ball C A, Blake J A, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nature Genetics 2000; 25:25-9.
13. Kanehisa M, Furumichi M, Tanabe M, et al. KEGG: new perspectives on genomes, pathways, diseases and drugs. Nucleic Acids Research 2017; 45:D353-D361.
14. Dennis G, Sherman B T, Hosack D A, et al. DAVID: Database for Annotation, Visualization, and Integrated Discovery. Genome Biology 2003; 4:P3.
15. Szklarczyk D, Franceschini A, Wyder S, et al. STRING v10: protein-protein interaction networks, integrated over the tree of life. Nucleic Acids Res 2015; 43: D447–D452.
16. Shannon, P. Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. Genome Research 2003; 13:2498-2504.
17. Tang Z , Li C , Kang B , et al. GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses . Nucleic Acids Research 2017; 45:W98–W102..
18. Hou G X,Liu P,Yang J , et al. Mining expression and prognosis of topoisomerase isoforms in non-small-cell lung cancer by using Oncomine and Kaplan–Meier plotter. PLoS ONE 2017; 12:e0174515.
19. Dwivedi P, Muench DE, Wagner M, et al. Phospho serine and threonine analysis of normal and mutated granulocyte colony stimulating factor receptors. Sci Data 2019; 6: 21.
20. Dwivedi P, Muench DE, Wagner M, et al. Time resolved quantitative phospho-tyrosine analysis reveals Bruton's Tyrosine kinase mediated signaling downstream of the mutated granulocyte-colony stimulating factor receptors. Leukemia 2019; 33: 75–87.
21. Zamanian-Azodi M, Rezaei Tavirani M, Rostami-Nejad M, et al. New Molecular Aspects of Cardiac Arrest; Promoting Cardiopulmonary Resuscitation Approaches. Emerg (Tehran) 2018; 6:e40.
22. Mou T , Zhu D , Wei X , et al. Identification and interaction analysis of key genes and microRNAs in hepatocellular carcinoma by bioinformatics analysis . World Journal of Surgical Oncology 2017; 15:63.

23. Liu J, Li H, Sun L, et al. Aberrantly methylated-differentially expressed genes and pathways in colorectal cancer. Cancer Cell International 2017; 17:75.
24. Sun C, Yuan Q, Wu D, et al. Identification of core genes and outcome in gastric cancer using bioinformatics analysis. Oncotarget 2017; 8:70271-70280.
25. Tang Z, Li C, Kang B, Gao G, et al. GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses. Nucleic Acids Res 2017; 45: W98–W102.
26. Rich J T, Neely J G, Paniello R C, et al. A practical guide to understanding Kaplan-Meier curves. Otolaryngology-Head and Neck Surgery 2010; 143:331-336.
27. Hung MS, Chen IC, Lung JH, et al. Epidermal Growth Factor Receptor Mutation Enhances Expression of Cadherin-5 in Lung Cancer Cells. PLoS One 2016;11:e0158395.
28. Mao X, Xue X, Wang L, et al. CDH5 is specifically activated in glioblastoma stemlike cells and contributes to vasculogenic mimicry induced by hypoxia. Neuro-Oncology 2013; 15:865.
29. Du J , Yang Q , Luo L , et al. C1qr and C1qrl redundantly regulate angiogenesis in zebrafish through controlling endothelial Cdh5. Biochemical and Biophysical Research Communications 2017; 483:482-487.
30. Mikus M, Drobin K, Gry M, et al. Elevated levels of circulating CDH5 and FABP1 in association with human drug-induced liver injury. Liver International 2016; 37:132-140.
31. Higuchi K , Inokuchi M , Takagi Y , et al. Cadherin 5 expression correlates with poor survival in human gastric cancer. Journal of Clinical Pathology 2017; 70: 217-221.
32. Jia Y, Qin T, Zhang X, et al. Effect of bevacizumab on the tight junction proteins of vascular endothelial cells. Am J Transl Res 2019; 11: 5546–5559.
33. Ma SC, Li Q, Peng JY, et al. Claudin-5 regulates blood-brain barrier permeability by modifying brain microvascular endothelial cell proliferation, migration, and adhesion to prevent lung cancer metastasis. CNS Neurosci Ther 2017; 23: 947–960.
34. Jia Y, Qin T, Zhang X, et al. Effect of bevacizumab on the tight junction proteins of vascular endothelial cells. Am J Transl Res 2019; 11: 5546–5559.
35. Menard C, Pfau M L, Hodes G E, et al. Social stress induces neurovascular pathology promoting depression. Nature Neuroscience 2017; 20:1752-1760.
36. Eskandar O, Fariba M M, Parima S, et al. Polymorphism of the CLDN5 gene and Schizophrenia in an Iranian Population. Iranian Journal of Public Health 2014; 43:79-83.
37. Du Y, Xu J T, Jin H N, et al. Increased cerebral expressions of MMPs, CLDN5, OCLN, ZO1 and AQPs are associated with brain edema following fatal heat stroke. Scientific Reports 2017; 7:1691.
38. Ma SC, Li Q, Peng JY, et al. CLDN5 affects lncRNAs acting as ceRNA dynamics contributing to regulating blood-brain barrier permeability in tumor brain metastasis. Oncol Rep 2018; 39: 1441–1453.
39. Ilhan-Mutlu A, Siehs C, Berghoff AS, et al. Expression profiling of angiogenesis-related genes in brain metastases of lung cancer and melanoma. Tumour Biol 2016; 37: 1173–1182.

40. Frezzetti D,Gallo M, Roma C, et al. Vascular Endothelial Growth Factor A Regulates the Secretion of Different Angiogenic Factors in Lung Cancer Cells. Journal of Cellular Physiology 2016; 231: 1514-1521.
41. Dunleavey J M, Xiao L, Thompson J, et al. Vascular channels formed by subpopulations of PECAM1+ melanoma cells, Nature Communications 2014; 5:5200.
42. Ohashi J, Naka I, Hananantachai H, et al. Association of PECAM1/CD31 polymorphisms with cerebral malaria. International Journal of Molecular Epidemiology and Genetics 2016; 7:87-94.
43. Guo F, Si C, Zhou M, et al. Decreased PECAM1-mediated TGF-β1 expression in the mid-secretory endometrium in women with recurrent implantation failure. Hum Reprod 2018; 33: 832–843.
44. Sun X, Huang S, Wang X, et al. CD300A promotes tumor progression by PECAM1, ADCY7 and AKT pathway in acute myeloid leukemia. Oncotarget 2018; 9: 27574–27584.
45. Musumeci G, Castorina A, Magro G, et al. Enhanced expression of CD31/platelet endothelial cell adhesion molecule 1 (PECAM1) correlates with hypoxia inducible factor-1 alpha (HIF-1α) in human glioblastoma multiforme. Exp Cell Res 2015; 339:407–416.

Tables

Table 1 **DEGs detected from three profile datasets**

DEGs	Genes name
up-regulated(21)	DNAH14 CCNB1 HMGB3 UCHL1 CHEK1 KIF18B AURKA MCM10 NUF2 DEPDC1 FAM83A IQGAP3 PHLDA2 TFAP2A HMMR EXO1 PCP4 NMU SIX1 NEK2 CENPF
down-regulated(116)	PKNOX2 SOX7 PPP1R14A ERG SYNPO2 GIMAP8 PCAT19 ACVRL1 GRK5 VGLL3 LTBP4 EMP2 SLCO2A1 GDF10 BCHE CD36 GPD1 NPR1 TBX2 SPTBN1 RASIP1 PTPRB QKI PIR-FIGF///FIGF ITGA8 MT1M TNNC1 ADRA1A MCEMP1 FHL1 THBD ABCA8 AOC3 ADRB1 NDRG2 SVEP1 TCF21 ASPA EDNRB SLIT3 SCN4B MYCT1 KANK3 STX11 MYH11 AGER SOX17 VWF ABI3BP CD93 TIE1 AGTRI FLI1 SH2D3C CLIC5 ADRB2 FGFR4 FHL5 SGCG PDK4 COL13A1 ANGPT1 DUOX1 EMCN MFAP4 PECAM1 OGN SCARA5 CLDN5 MAOB ATP1A2 IGSF10 SCGB1A1 CDO1 CA4 SDPR CLIC3 S1PR1 LYVE1 ADAMTS8 LEPROT//LEPR SPOCK2 AKAP12 HSPA12B ROBO4 CALCRL CAV1 JAM2 FOXF1 DST FGD5 RHOJ FMO2 SHROOM4 SFTPC TTN TGFBR3 HHIP ADH1B FABP4 GPC3 FAM107A PGM5 GPX3 MARCO SEMA5A RAMP2 KIAA1462 EPAS1 SLIT2 ADAMTSL3 CLDN18 C2orf40 CDH5

Table 2 **Gene ontology analysis of differentially expressed genes in lung cancer**

Expression	Category	Term	Count	%	p-Value	FDR
up-regulated	GOTERM_BP_FAT	GO:0000279 M phase	7	33.33	1.95E-06	0.0027
	GOTERM_BP_FAT	GO:0022403 cell cycle phase	7	33.33	7.36E-06	0.0101
	GOTERM_BP_FAT	GO:0022402 cell cycle process	7	33.33	4.31E-05	0.0593
	GOTERM_BP_FAT	GO:0000278 mitotic cell cycle	6	28.57	7.03E-05	0.0966
	GOTERM_BP_FAT	GO:0007067 mitosis	5	23.81	1.37E-04	0.1885
	GOTERM_BP_FAT	GO:0000280 nuclear division	5	23.81	1.37E-04	0.1885
	GOTERM_BP_FAT	GO:0000087 M phase of mitotic cell cycle	5	23.81	1.47E-04	0.2021
	GOTERM_BP_FAT	GO:0048285 organelle fission	5	23.81	1.60E-04	0.2200
	GOTERM_BP_FAT	GO:0007017 microtubule-based process	5	23.81	2.35E-04	0.3224
	GOTERM_BP_FAT	GO:0007049 cell cycle	7	33.33	2.50E-04	0.3439
	GOTERM_CC_FAT	GO:0015630 microtubule cytoskeleton	7	33.33	2.19E-05	0.0218
	GOTERM_CC_FAT	GO:0043228 non-membrane-bounded organelle	11	52.38	1.28E-04	0.1277
	GOTERM_CC_FAT	GO:0043232 intracellular non-membrane-bounded c	11	52.38	1.28E-04	0.1277
	GOTERM_CC_FAT	GO:0000793 condensed chromosome	4	19.05	4.18E-04	0.4153
	GOTERM_CC_FAT	GO:0044430 cytoskeletal part	7	33.33	4.68E-04	0.4649
down-regulated	GOTERM_BP_FAT	GO:0001944 vasculature development	15	13.39	1.86E-09	0.0000
	GOTERM_BP_FAT	GO:0001568 blood vessel development	14	12.50	1.35E-08	0.0000
	GOTERM_BP_FAT	GO:0048514 blood vessel morphogenesis	12	10.71	2.28E-07	0.0004
	GOTERM_BP_FAT	GO:0007155 cell adhesion	19	16.96	1.26E-06	0.0021
	GOTERM_BP_FAT	GO:0022610 biological adhesion	19	16.96	1.29E-06	0.0021
	GOTERM_BP_FAT	GO:0003013 circulatory system process	10	8.93	5.63E-06	0.0092
	GOTERM_BP_FAT	GO:0008015 blood circulation	10	8.93	5.63E-06	0.0092
	GOTERM_BP_FAT	GO:0001525 angiogenesis	9	8.04	8.54E-06	0.0139
	GOTERM_BP_FAT	GO:0032101 regulation of response to external stim	9	8.04	1.44E-05	0.0235
	GOTERM_BP_FAT	GO:0003018 vascular process in circulatory system	6	5.36	4.98E-05	0.0810
	GOTERM_BP_FAT	GO:0044057 regulation of system process	11	9.82	5.54E-05	0.0900
	GOTERM_BP_FAT	GO:0008217 regulation of blood pressure	7	6.25	6.76E-05	0.1097
	GOTERM_BP_FAT	GO:0042127 regulation of cell proliferation	17	15.18	9.48E-05	0.1540
	GOTERM_BP_FAT	GO:0042312 regulation of vasodilation	4	3.57	3.40E-04	0.5511
	GOTERM_BP_FAT	GO:0035150 regulation of tube size	5	4.46	4.93E-04	0.7988
	GOTERM_BP_FAT	GO:0050880 regulation of blood vessel size	5	4.46	4.93E-04	0.7988
	GOTERM_BP_FAT	GO:0048738 cardiac muscle tissue development	5	4.46	6.52E-04	1.0537
	GOTERM_BP_FAT	GO:0007517 muscle organ development	8	7.14	6.61E-04	1.0681
	GOTERM_BP_FAT	GO:0035295 tube development	8	7.14	8.45E-04	1.3642
	GOTERM_CC_FAT	GO:0044459 plasma membrane part	46	41.07	1.56E-11	0.0000
	GOTERM_CC_FAT	GO:0005886 plasma membrane	56	50.00	1.29E-08	0.0000
	GOTERM_CC_FAT	GO:0005887 integral to plasma membrane	29	25.89	2.47E-08	0.0000
	GOTERM_CC_FAT	GO:0031226 intrinsic to plasma membrane	29	25.89	4.01E-08	0.0000
	GOTERM_CC_FAT	GO:0044421 extracellular region part	21	18.75	2.40E-05	0.0298
	GOTERM_CC_FAT	GO:0009986 cell surface	12	10.71	5.80E-05	0.0722
	GOTERM_CC_FAT	GO:0015629 actin cytoskeleton	10	8.93	1.86E-04	0.2311
	GOTERM_CC_FAT	GO:0005578 proteinaceous extracellular matrix	10	8.93	6.62E-04	0.8215
	GOTERM_CC_FAT	GO:0009925 basal plasma membrane	4	3.57	9.42E-04	1.1661
	GOTERM_MF_FAT	GO:0050431 transforming growth factor beta bindin	4	3.57	3.99E-05	0.0533
	GOTERM_MF_FAT	GO:0051015 actin filament binding	5	4.46	5.30E-04	0.7074

Table 3 **DEGs via KEGG pathway analysis in lung cancer**

Expression	Term	Count	%	P-Value	Genes
up-regulated	hsa04115:p53 signaling pathway	2	9.52	7.76E-02	CCNB1, CHEK1
down-regulated	hsa04270:Vascular smooth muscle contraction	7	6.25	4.56E-04	AGTR1, RAMP2, MYH11, NPR1, ADRA1A, CALCRL, PPP1R14A
	hsa04514:Cell adhesion molecules (CAMs)	6	5.36	6.38E-03	CLDN18, ITGA8, PECAM1, CLDN5, JAM2, CDH5
	hsa05414:Dilated cardiomyopathy	5	4.46	9.17E-03	ADRB1, SGCG, TNNC1, ITGA8, TTN
	hsa04020:Calcium signaling pathway	6	5.36	2.04E-02	AGTR1, EDNRB, ADRB2, ADRB1, TNNC1, ADRA1A
	hsa04670:Leukocyte transendothelial migration	5	4.46	2.12E-02	CLDN18, PECAM1, CLDN5, JAM2, CDH5
	hsa04080:Neuroactive ligand-receptor interaction	7	6.25	2.65E-02	AGTR1, EDNRB, ADRB2, ADRB1, S1PR1, ADRA1A, CALCRL
	hsa05410:Hypertrophic cardiomyopathy (HCM)	4	3.57	4.09E-02	SGCG, TNNC1, ITGA8, TTN
	hsa00350:Tyrosine metabolism	3	2.68	5.96E-02	MAOB, ADH1B, AOC3

Table 4 Vadidation of 27 genes via GEPIA

Category	Genes
Genes with high expressed in lung cancer	NEK2 MCM10 CCNB1 CENPF AURKA CHEK1 EXO1 HMMR NUF2 DEPDC1 KIF18B
Genes with low expressed in lung cancer	CDH5 VWF CAV1 PECAM1 CLDN5 LYVE1 EMCN ADRA1A PTPRB GRK5 TIE1 ROBO4 EDNRB AGTR1 ANGPT1
Genes with low expressed in lung cancer(no significantly statistical difference)	NUM

Table 5 The prognosis of the 26 key genes

Category	Genes
Genes with significantly worse survival	AURKA CCNB1 CENPF CHEK1 DEPDC1 EXO1 HMMR KIF18B MCM10 NEK2 NUF2 AGTR1 ANGPT1 CAV1 CDH5 CLDN5 EDNRB EMCN GRK5 NMU PECAM1 PTPRB ROBO4 VWF
Genes without significantly worse survival	LYVE1 ADRA1A TIE1

Table 6 Re-analysis of 23 selected genes via KEGG pathway enrichment

Pathway ID	Name	Count	%	p-Value	Genes
hsa04670	Leukocyte transendothelial migration	3	13.04	3.52E-02	PECAM1, CLDN5, CDH5
hsa04514	Cell adhesion molecules (CAMs)	3	13.04	4.33E-02	PECAM1, CLDN5, CDH5

Figures

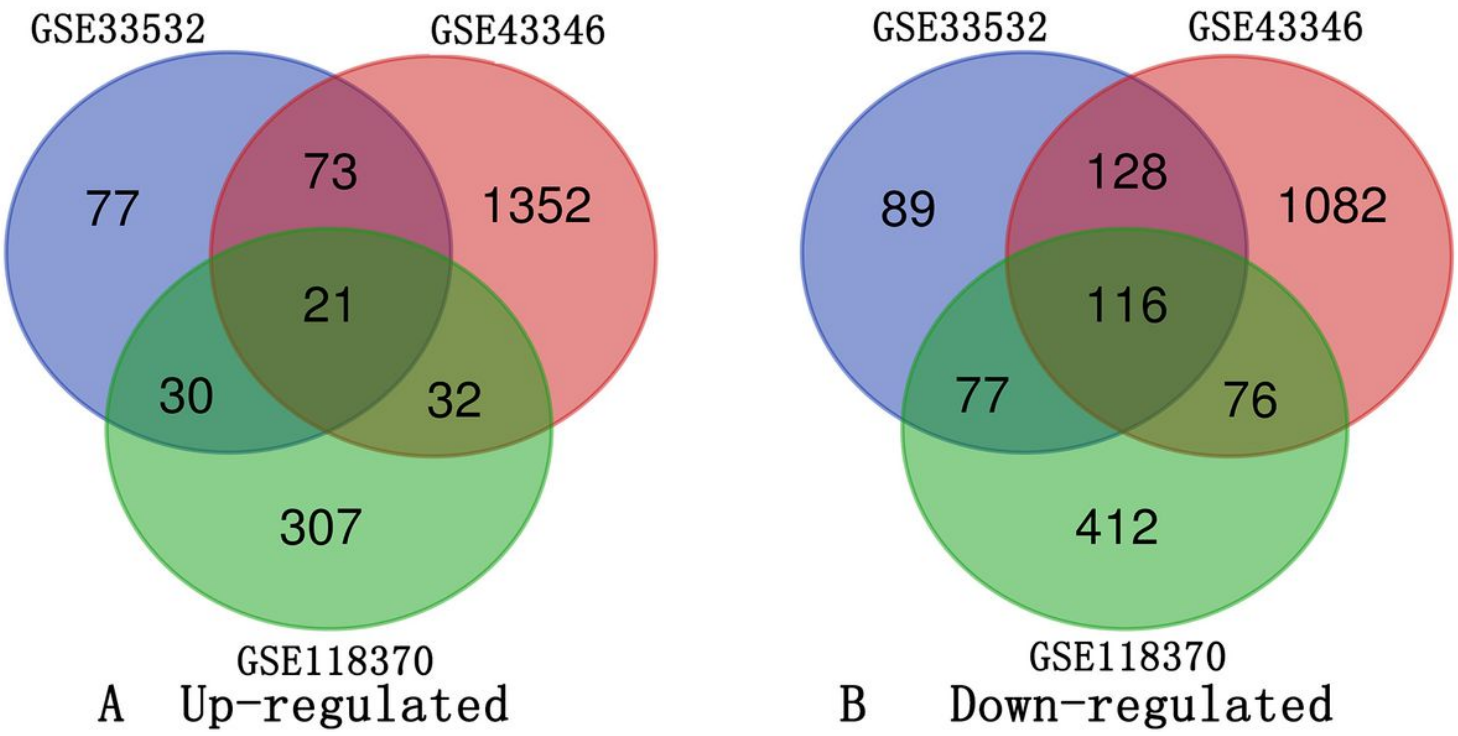

Figure 1

Identification of 137 common DEGs in the three datasets (GSE 33532, GSE 43346 and GSE 118370) through Venn diagrams software. Different color meant different datasets. 21 DEGs were up-regulated in the three datasets (logFC>2)(A). 116 DEGs were down-regulated in three datasets (logFC < -2)(B)

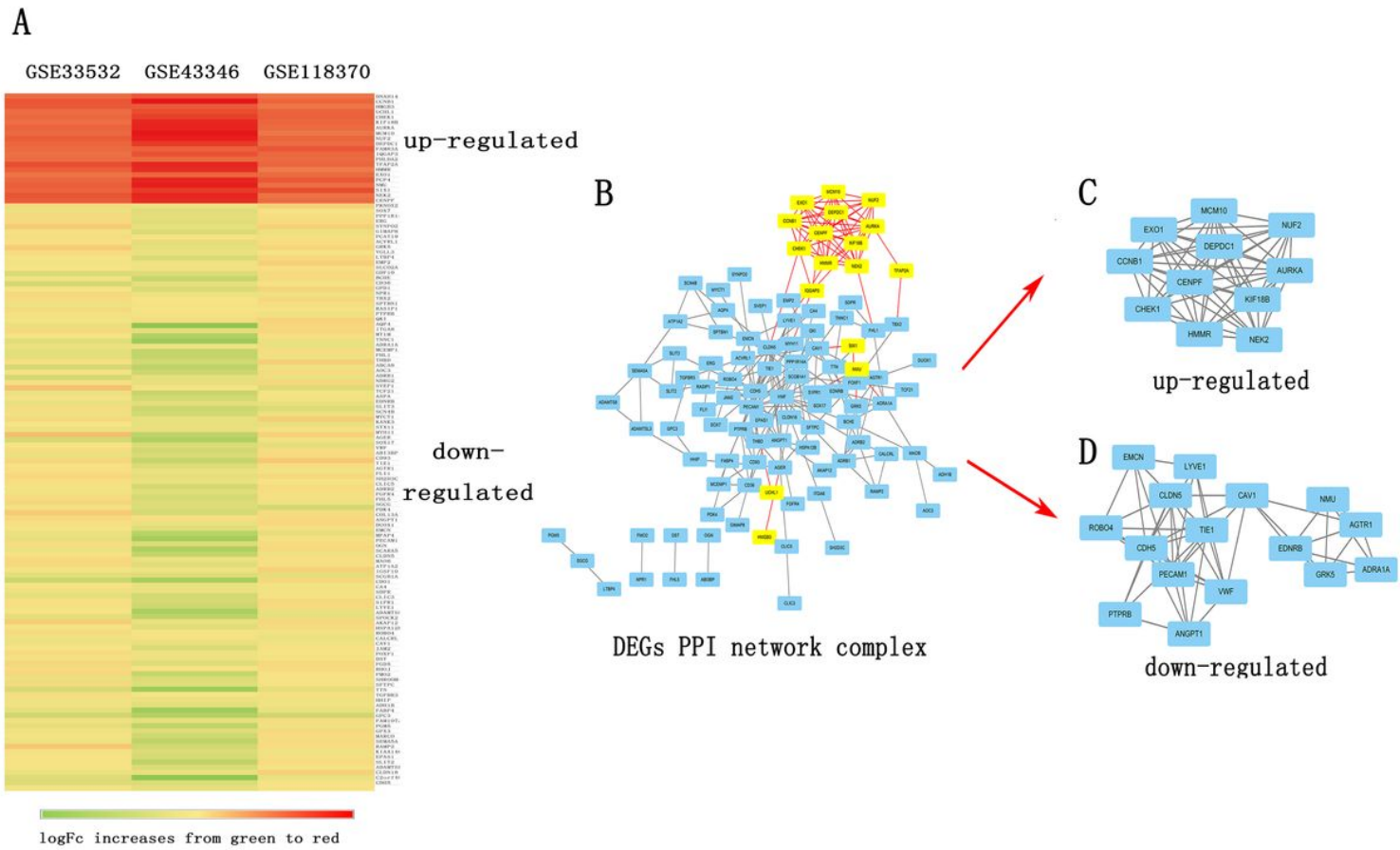

Figure 2

DEGs PPI network was constructed by STRING online database and Module analysis. Heatmap showed the expression of all 137 DEGs (A). There were a total of 21 DEGs in the DEGs PPI network complex. The rectangles meant proteins; the edges meant the interaction of proteins; blue rectangles meant down-regulated DEGs and yellow rectangles meant up-regulated DEGs (B). Module analysis via Cytoscape software (degree cut off=2, node score cut off=0.2, k-core= 2, and max. Depth= 100) (C and D).

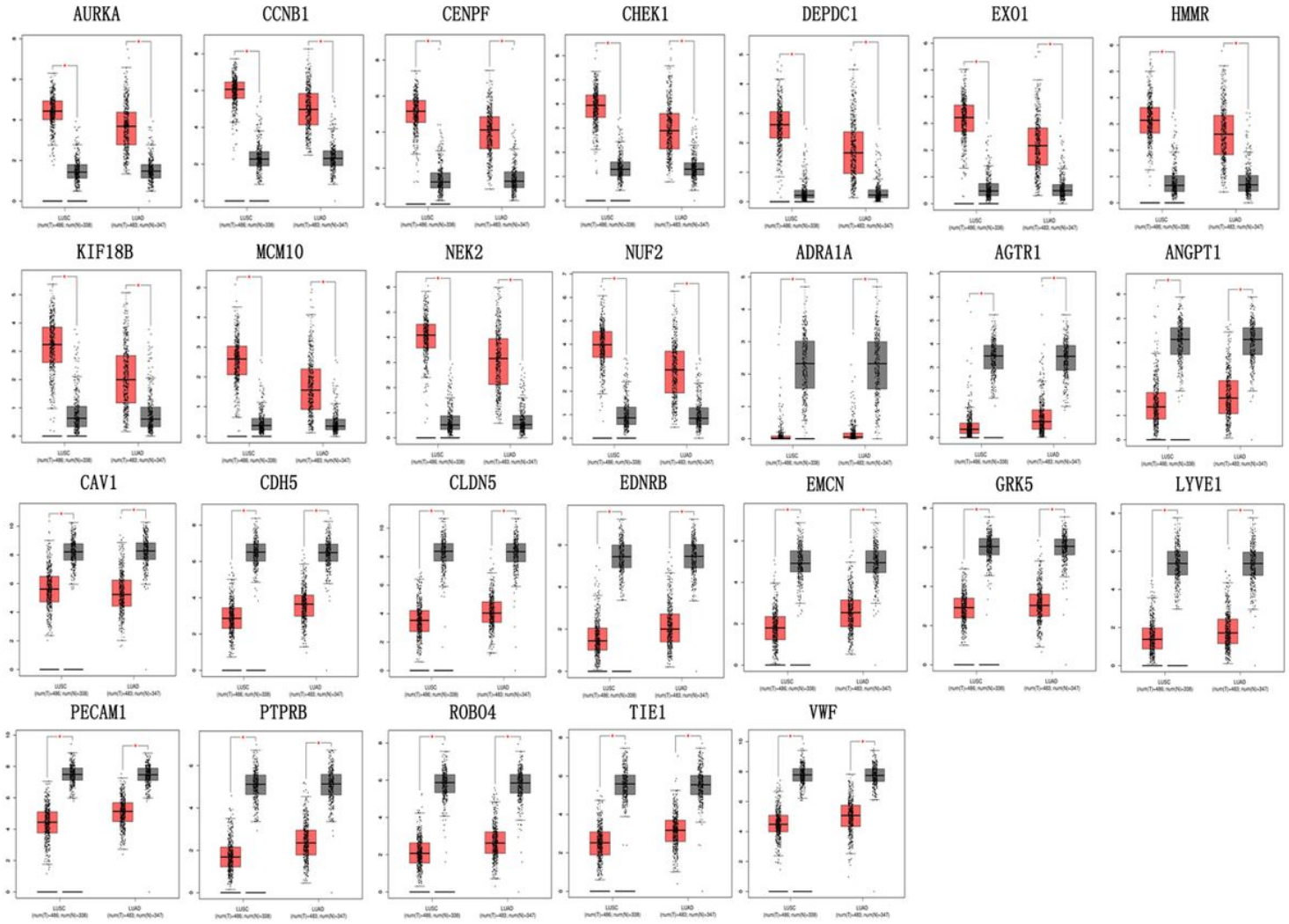

Figure 3

Significantly expressed 27 core genes in lung cancer samples. To further identify the level of genes expression between lung cancer samples and normal lung samples, 27 core genes were analyzed by GEPIA website. 26 of 27 genes had significant expression level in lung cancer specimen compared to normal specimen ($p < 0.05$). Red color means tumor tissues and grey color means normal tissues.

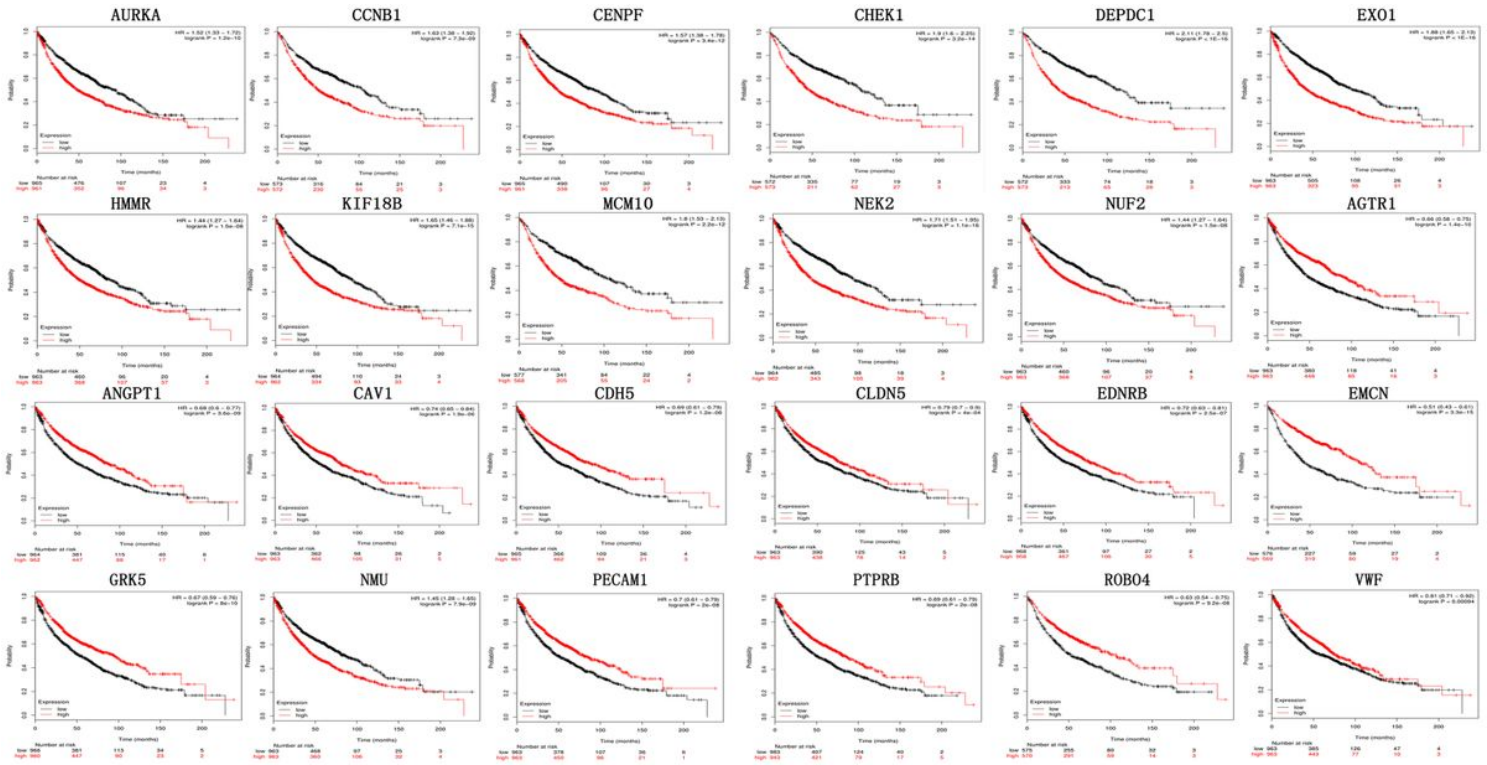

Figure 4

The prognostic information of the 26 core genes. Kaplan-meier plotter online tool was used to identify the prognositc information of the 26core genes and 23 of 26 genes had a significantly worse survival rate ($p < 0.05$).

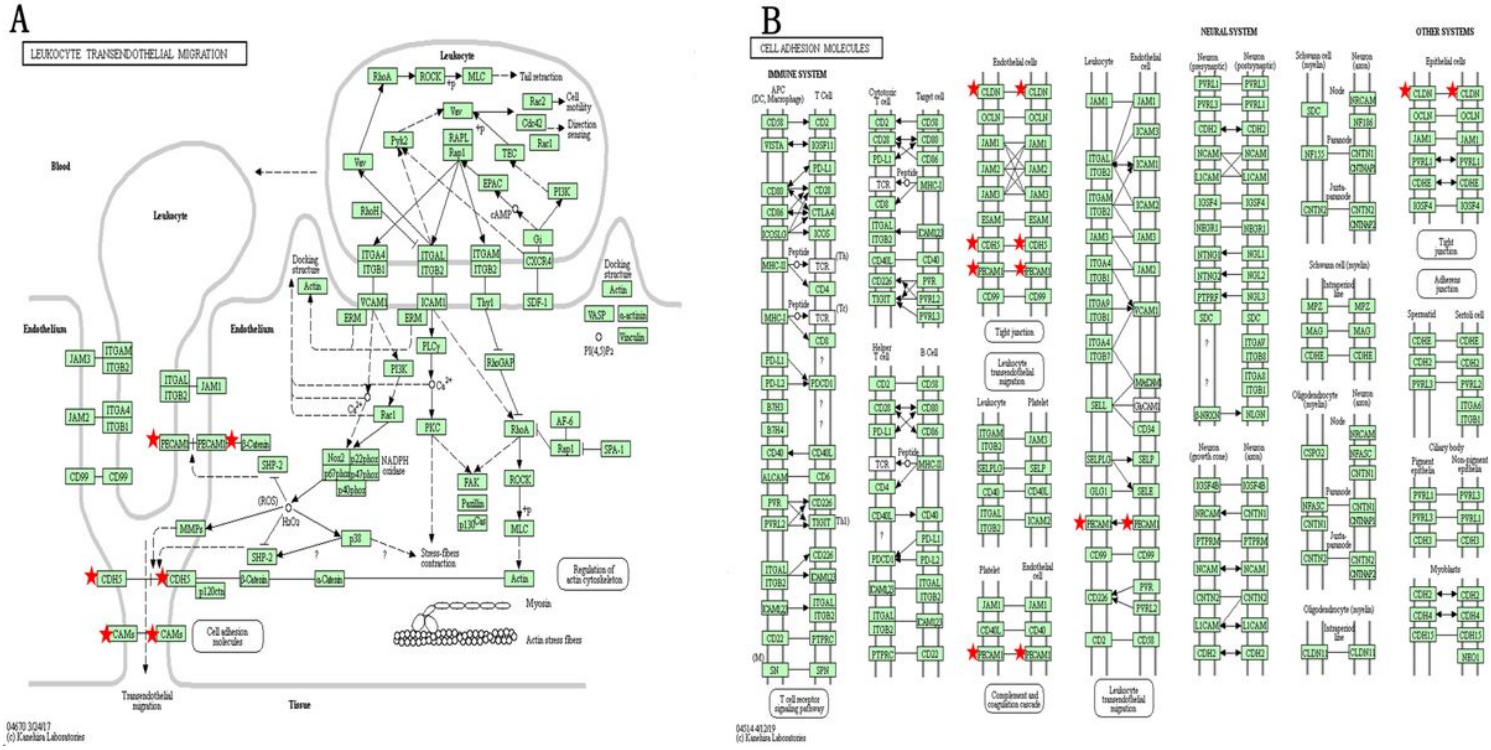

Figure 5

Re-analysis of 23 selected genes. Twenty-three genes in lung cancer samples with poor prognosis were re-analyzed by KEGG pathway enrichment(A). Three genes (CDH5, CLDN5, PECAM1) were significantly enriched in leukocyte trans-endothelial migration and cell adhesion molecules.(B)