

Integrative Bioinformatic and Functional Analyses of Potential Genes in Female Patients with Non-Small-Cell Lung Cancer

Xianghai Wang (✉ wxhwxpcyy@163.com)

Yijishan Hospital of Wannan Medical College

Hanli Wang

Yijishan Hospital of Wannan Medical College

Xingwu Chen

Yijishan Hospital of Wannan Medical College

Tangfeng Lv

Nanjing University School of Medicine

Research Article

Keywords: NSCLC, females, bioinformatic analysis, DEGs

Posted Date: April 9th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-386332/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Integrative Bioinformatic and Functional Analyses of Potential Genes in Female Patients with Non-Small-Cell Lung Cancer

Xianghai Wang¹ Hanli Wang¹ Xingwu Chen¹ Tangfeng Lv²

1 Department of Respiratory Medicine, Yijishan Hospital of Wannan Medical College, Wuhu, Anhui 241001, People's Republic of China; 2

Department of Respiratory Medicine, Jinling Hospital, Nanjing University School of Medicine, Nanjing, Jiangsu 210002, People's Republic of China

Correspondence: Xianghai Wang

Department of Respiratory Medicine, Yijishan Hospital of Wannan Medical College, 2 West Zheshan Road, Wuhu, Anhui 241001, People's Republic of China

Email: wxhwxpcyy@163.com

[Abstract] Objective: To explore the pathogenesis and prognostic biomarkers of non-small-cell lung cancer (NSCLC) in female patients via a bioinformatic analysis and functional prediction of potential NSCLC-associated genes in females. **Methods:** Data for female patients with NSCLC were downloaded from the Gene Expression Omnibus (GEO) database, and differentially expressed genes (DEGs) were identified using GEO2R. The DAVID online database was used to perform Gene Ontology (GO) and Kyoto Encyclopaedia of Genes and Genomes (KEGG) analyses, and STRING online software was used to

perform protein-protein interaction (PPI) analyses. Next, the plug-in (M-CODE) was used to screen the key DEGs; the Oncomine database was analysed for IL6, EGF and MMP9 expression in NSCLC tissues and normal lung tissues and the Kaplan-Meier plotter was used to perform prognostic analyses of key DEGs. Finally, RT-PCR was used to verify the expression of key DEG in NSCLC cells. **Results:** A total of 500 DEGs were screened, and the functional and pathway enrichment analysis showed that these genes were mainly involved in cell proliferation, cell migration, and vasculature development regulation. The KEGG analysis showed that the pathways were primarily related to ECM-receptor interactions, protein digestion and absorption, and leukocyte transendothelial migration signalling. Three key DEGs were obtained by the PPI network analysis: IL6, EGF and MMP9. The expression of IL6 was low in NSCLC tissues, while that of EGF and MMP9 was high. IL6 and EGF may be biomarkers for predicting the prognosis of female patients with NSCLC. Compared with human bronchial epithelial cell line (16HBE), EGF and MMP-9 were high expressed in NSCLC cells.

Conclusion: In female patients, IL6, EGF and MMP9 may be research targets for characterizing the pathogenesis of NSCLC, and IL6 and EGF may be biomarkers for predicting the prognosis of this cancer.

[Keywords] NSCLC; females; bioinformatic analysis; DEGs

Introduction

Lung cancer has become a major public health problem worldwide, and smoking is considered an important cause[1-2]. In the last 10 years, the proportion of lung cancer in women has gradually increased compared with that in men, and the vast majority of female lung cancer patients have no history of smoking. Studies have shown that 15% of male and 53% of female lung cancer cases are not related to smoking, and NSCLC has a higher incidence and worse prognosis in women[3-4]. At present, female NSCLC patients represent a special and large group of cancer patients. Environmental factors and genetic susceptibility may be the main factors leading to the risk of female NSCLC[5]. In this study, we downloaded a gene expression microarray of female patients with NSCLC from the Gene Expression Omnibus (GEO) and used GEO2R software to compare NSCLC tissue with normal lung tissue to screen out DEGs and evaluate their clinical value with the severity and prognosis of the disease. The purpose of this study was to further investigate the molecular mechanism of NSCLC in females and its biomarkers for early diagnosis, precise treatment and prognosis.

1 Materials and methods

1.1 Chip data sources and DEG screening

GSE19804 and GSE31210 chip data sets were downloaded from the GEO database (<https://www.ncbi.nlm.nih.gov/geo/>). Both chip platforms are GPL570. After analysing NSCLC tissues and normal lung tissues with

GEO2R software in the GEO database, the DEGs of NSCLC tissues and normal lung tissues were obtained and screened. The array data of GSE19804 included 60 female NSCLC tissue samples and 60 healthy female lung tissues[6]. GSE31210 contained 246 samples, consisting of 121 female NSCLC samples and 9 normal female lung tissues[7]. Screening criteria: $\log_{2}FC > 1$ or $\log_{2}FC < -1$.

1.2 Enrichment analysis of GO and KEGG

GO and KEGG analyses and functional annotation of DEGs were carried out using the David online database. The main biological processes and tumor-related signalling pathways associated with the DEGs were analysed to obtain a more comprehensive understanding of their mechanisms and functions.

1.3 Screening of key DEGs and analysis of the relationship between DEG expression level and prognosis of female NSCLC patients

String (<https://string-db.org/>) is a powerful online analysis website for protein-protein interaction networks (PPIs). A PPI network of key DEGs was constructed using this website, and the screening condition was a combined score > 0.4 . The M-CODE plug-in was used to screen out the most critical genes in the PPI network. The log rank test and Cox proportional hazards regression method were used to compare the overall survival (OS) of the two groups to verify the relationship between the expression of these key genes and the prognosis of female NSCLC

patients. The statistical analyses were performed using R software (version 3.3.2), with the significance level set as a two-sided P value of < 0.05.

1.4 Expression profile of key DEGs in tissues from the GEO and Oncomine datasets

The GSE19804 and GSE31210 datasets were downloaded from GEO. The relative expression level of key DEGs was extracted from each cohort. An unpaired t test was used to evaluate the differential expression of key DEGs between NSCLC tissues and adjacent normal tissues. The statistical analyses were performed using R software (version 3.3.2) with the significance level set as a two-sided P value of < 0.05.

The relative expression levels of key DEGs in different NSCLC datasets were analysed by searching the Oncomine bioinformatics gene database (<https://www.oncomine.org/>). The specific screening conditions were as follows. (1) gene: enter key DEG name; (2) cancer type: select cancer vs normal analysis; (3) cancer type: select NSCLC; (4) demographics: select female; and (5) analysis type: set P-value to 0.05, fold change to 2, and gene rank to 10%.

1.5 Cell lines

The 16HBE, lung adenocarcinoma cell line (PC9) and lung squamous cell carcinoma (SK-MES-1) were purchased from the Institute of Biochemistry and Cell Biology of the Chinese Academy of Sciences.

16HBE and PC9 cells were maintained in RPMI-1640 medium (Gibco Company, Grand Island, NY). SK-MES-1 cell was maintained in DMEM medium (Gibco Company, Grand Island, NY). The media was supplemented with 10% foetal bovine serum (FBS). The cells were grown in a humidified atmosphere at 37°C with 5% CO₂.

1.6 RNA isolation, complementary DNA (cDNA) synthesis and quantitative reverse transcriptase polymerase chain reaction (qRT-PCR) analyses

Total RNA was extracted from cultured cells using TRIzol reagent (Invitrogen). cDNA was synthesized using a Reverse Transcription Kit (TaKaRa). We performed qRT-PCR using a SYBR Premix Ex Taq II kit (TaKaRa), and the target gene expression values were normalized to the expression of glyceraldehyde-3-phosphate dehydrogenase (GAPDH). The results were analysed based on the comparative cycle threshold (CT) ($2^{-\Delta\Delta CT}$) methods. The primer sequences used in this study were listed in Supplement Table S1.

2 Results

2.1 Five hundred common DEGs were found to be associated with NSCLC in females

The results showed that GSE19804 contained 1160 DEGs and GSE31210 contained 780 DEGs. After removing duplicate and incomplete data, a total of 500 common DEGs were found (Figure 1).

2.2 GO and KEGG analyses of the biological functions of 500 DEGs.

A cluster profiler was used to analyse all 500 DEGs. In the biological process category, the DEGs were mainly enriched in endothelial cell proliferation, amoeboid-type cell migration, urogenital system development, vasculature development regulation, and epithelial cell proliferation regulation. In terms of cell components, the DEGs were mainly enriched in glycosaminoglycan binding, extracellular matrix structural construct, and amide binding. In cell components, the DEGs were mainly enriched in glycosaminoglycan binding, extracellular matrix structural construction, and amide binding (Figure 2A-C). The results of the KEGG enrichment analysis showed that the DEGs were mainly enriched in ECM receptor interaction, malaria, protein digestion and absorption, and leukocyte transverse migration (Figure 2D).

2.3 PPI network was constructed, and 3 key DEGs were screened out.

Five hundred common DEGs were imported into STRING for analysis to obtain a PPI network. The PPI network was composed of 482 nodal proteins and 1678 interactions. Finally, three genes located in key positions were screened out: interleukin-6 (IL6), epidermal growth factor (EGF) and matrix metalloproteinase 9 (MMP9). The corresponding degrees of IL6, EGF and MMP9 were 75, 58 and 42, respectively (Figure 3).

2.4 Expression of key DEGs and their relationship with prognosis

We used the GEPIA tool to verify the expression of three key genes in NSCLC tissues and normal lung tissues based on the GEO database (Figure 4A-F) and further verified the correlation between the expression levels of the three key genes and the prognosis of female NSCLC patients using GEO31210 chip data. The survival analysis showed that IL6 and EGF were associated with the prognosis of NSCLC in female patients (Figure 5A-C).

2.5 Oncomine database analysis of IL6, EGF and MMP9 expression in NSCLC tissues and normal lung tissues

An analysis of eight datasets found that the expression of IL6 in NSCLC tissues was lower than that in normal lung tissues, and the expression of EGF and MMP9 in NSCLC tissues was higher than that in normal lung tissues (Figure 6A-C).

2.6 Expression of key DEG in NSCLC cells

The expression of three key genes in NSCLC cells were assessed using RT-qPCR, the expression of IL6 in NSCLC cells was lower than that in 16HBE cell line, while the expression of EGF and MMP9 in NSCLC cells was higher than that in 16HBE cell line (Figure 7A-C).

3 Discussion

With the continuous development of medical care, the proportion of female patients with LAD among new lung cancer cases is increasing significantly[8]. At present, the specific molecular mechanism of NSCLC

in females has not been fully elucidated, although it has been shown to involve the abnormal expression, inactivation, and mutation of many genes and the involvement of a variety of related signalling pathways[9]. With the development of second-generation sequencing technology, the use of bioinformatic methods to analyse the treatment targets of female patients with NSCLC is a current research hotspot[10]. GEO2R software was used to screen the GSE19804 and GSE31210 datasets of tissue microarray expression of female NSCLC patients in the GEO database. A total of 500 DEGs were screened.

Then, the functional and pathway enrichment analysis of these DEGs showed that they were mainly involved in cell proliferation, cell migration, and vasculature development regulation. The KEGG analysis showed that it was mainly related to ECM-receptor interaction, protein digestion and absorption, leukocyte transendothelial migration and other signalling pathways, and these findings facilitated a further exploration of the mechanisms. Three key genes (IL6, EGF and MM9) closely related to the occurrence and development of NSCLC in females were screened, which provided potential targets for the follow-up study of the treatment of NSCLC.

IL-6 can regulate tumor activity in two ways. Its main cancer-promoting effect is to produce Th17 cells and inhibit dendritic cell and CD8⁺ T cell activity. IL-6 suppresses cancer by inhibiting regulatory

cells[11]. The intricate interaction between IL6-mediated tumor immunosuppression and metabolic disorders was also observed in colorectal and pancreatic cancer[12]. In this study, we found that the expression of IL-6 in NSCLC tissues and serum was lower than that in normal tissues and serum. Therefore, according to our preliminary research results, we speculate that IL6 exhibits a two-way regulatory role in the process of tumour immunity. The low expression level of IL-6 in the serum of patients with NSCLC may be one of the important factors promoting the occurrence and development of NSCLC, although its specific mechanism remains to be clarified.

Epidermal growth factor (EGF) is a low molecular weight peptide chain composed of 53 amino acid residues and a strong mitogenic agent. It can promote the growth of tumor cells and induce tumor cell invasion and is closely related to the prognosis of a variety of cancer patients, such as liver cancer, gastric cancer and breast cancer[13-14]. Matrix metalloproteinases (e.g., MMP-9) are the most important proteolytic enzymes that degrade the extracellular matrix. MMP-9 is an important matrix metalloproteinase in the MMP family. It is a Zn^{2+} -dependent extracellular proteolytic enzyme that can degrade extracellular matrix and promote the invasion and metastasis of tumor cells[15].

In conclusion, based on big data, three key DEGs and their biological functions in female NCSLC patients were obtained based on a

bioinformatic analysis. An analysis of the GEO and Oncomine databases found that the expression of IL6 in NSCLC tissues was lower than that in normal lung tissues and the expression of EGF and MMP9 in NSCLC tissues was higher than that in normal lung tissues. Additionally, the expression of IL6, EGF and MMP9 in NSCLC cells were verified by RT-qPCR, which further supported the bioinformatics results. Among them, IL6 and EGF are associated with the prognosis of NSCLC in females, which provides a potential target for follow-up studies on the prognosis of NSCLC in females. At present, there are few studies on the correlation between IL-6 and the occurrence and development of NSCLC. Although big data can reveal correlations, big data are not equal to evidence; thus, further basic and clinical studies are needed.

Acknowledgements

We sincerely thank the researchers for providing their GEO and Oncomine databases information online, it is our pleasure to acknowledge their contributions.

Authors' contributions

XHW contributed to the study design, XHW, HLW, XWC, TFL conducted the literature search. XHW carried out the Cell experiments. XHW, HLW and TFL performed data analysis and drafted. All authors have read and approved the manuscript.

Funding

The study was sponsored by National Natural Science Foundation of China (grant no. 81772500), Yijishan Hospital of Wannan Medical College Introduces Talent Foundation and Excellent Young Talents Support Programme in Colleges and Universities.

Availability of data and materials

The datasets analysed during the current study are available in the Gene Expression Omnibus repository and Oncomine repository, <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE19804>, <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE31210>, <https://www.oncomine.org/resource/login.html>.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no conflicts of interest.

Authors' information

1 Department of Respiratory Medicine, Yijishan Hospital of Wannan Medical College, 2 West Zheshan Road, Wuhu, Anhui 241001, China. 2 Department of Respiratory Medicine, Jinling Hospital, Nanjing University School of Medicine, Nanjing, Jiangsu 210002, China

1 Department of Pulmonary and Critical Care Medicine, Haining People's Hospital, Jiaxing 314400, China. 2 Department of Respiratory and Critical Care Medicine, Changhai Hospital, Naval Medical University (Second Military Medical University), No. 168 Changhai Road, Yangpu District, Shanghai 200433, China.

References

- [1] LEVY D T, MEZA R, ZHANG YA, et al. Gauging the effect of U.S. tobacco control policies from 1965 through 2014 using SimSmoke. *Am J Prev Med*, 2016, 50(4): 535-542.
- [2] SAUER A G, SIEGEL R L, JEMAL A, et al. Updated review of prevalence of major risk factors and use of screening tests for cancer in the United States. *Cancer Epidemiol Biomarkers Prev*, 2017, 26(8): 1192-1208.
- [3] SUN S, SCHILLER J H, GAZDAR A F. Lung cancer in never smokers-a different disease. *Nat Rev Cancer*, 2007, 7(10): 778-790.
- [4] KIM J H, PARK K, YIM S H, et al. Genome-wide association study of lung cancer in Korean non-smoking women. *J Korean Med Sci*, 2013, 28(6): 840-847.
- [5] HUBER R M. Is lung cancer in never-smokers a different disease?-- Back to the figures. *J Thorac Oncol*, 2007, 2(9): 787-788.
- [6] Lu TP, Tsai MH, Lee JM et al: Identification of a novel biomarker

SEMA5A for non-small cell lung carcinoma in nonsmoking women. *Cancer Epidemiol Biomarkers Prev*, 2010; 19(10): 2590–2597.

[7] Yamauchi M, Yamaguchi R, Nakata A et al: Epidermal growth factor receptor tyrosine kinase defines critical prognostic genes of stage I lung adenocarcinoma. *PLoS One*, 2012; 7(9): e43923.

[8] STILES B M, RAHOUMAM, HUSSEIN M K, et al. Never smokers with resected lung cancer: different demographics, similar survival. *Eur J Cardiothorac Surg*, 2018, 53(4): 842-848.

[9] CHOI J R, PARK S Y, NOH O K, et al. Gene mutation discovery research of nonsmoking lung cancer patients due to indoor radon exposure. *Ann Occup Environ Med*, 2016, 28: 13

[10] ZHAO J G, WANG J F, FENG J F, et al. HHIP overexpression inhibits the proliferation, migration and invasion of non-small cell lung cancer. *PLoS One*, 2019, 14(11): e0225755.

[11]H. Tsukamoto, K. Fujieda, S. Senju, et al. Immune-suppressive effects of interleukin-6 on T-cell-mediated antitumour immunity, *Cancer Sci*. 2018, 109(3): 523-530.

[12] T. R. Flint, T. Janowitz, C. M. Connell, et al. Tumor-Induced IL-6 Reprograms Host Metabolism to Suppress Antitumour Immunity, *Cell Metab*. 2016, 24(5): 672-684.

[13]JIANG W, TIAN W, IJAZ M, et al. Inhibition of EGF induced migration and invasion by sulfated polysaccharide of *Sepiella maindroni*

ink via the suppression of EGFR/Akt/p38/MAPK/MMP-2 signalling pathway in KB cells. Biomed Pharmacother,2017(11):95-102.

[14]KIM J,KONG J,CHANG H,et al.EGF induces epithelial-mesenchymal transition through phospho-Smad2/3-Snail signalling pathway in breast cancer cells. Oncotarget,2016(51):85021-85032.

[15] Ni Y, Ye X, Wan C, et al.Percutaneous microwave ablation (MWA) increased the serum levels of VEGF and MMP-9 in stage I non-small-cell lung cancer (NSCLC) .Int J Hyperthermia, 2017,33(4):435-439.

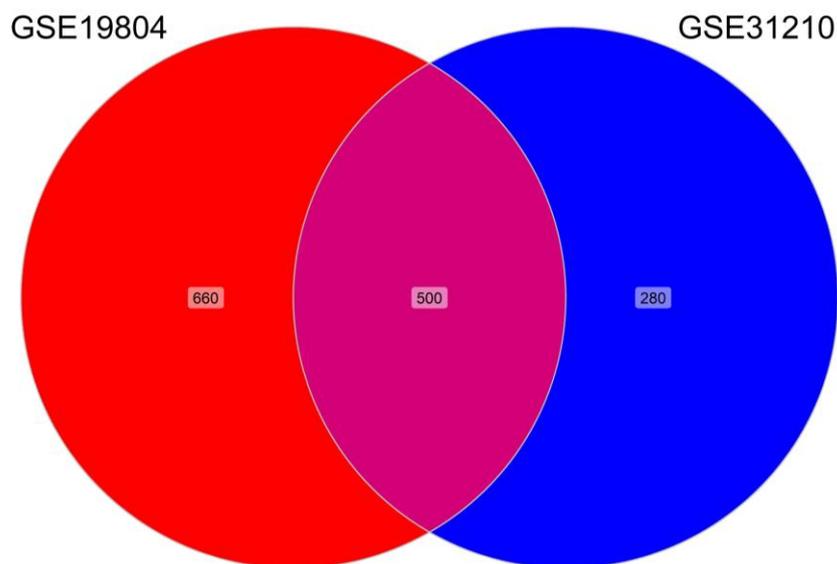


Figure 1 Five hundred common DEGs were screened from the GSE19804 and GSE31210 datasets.

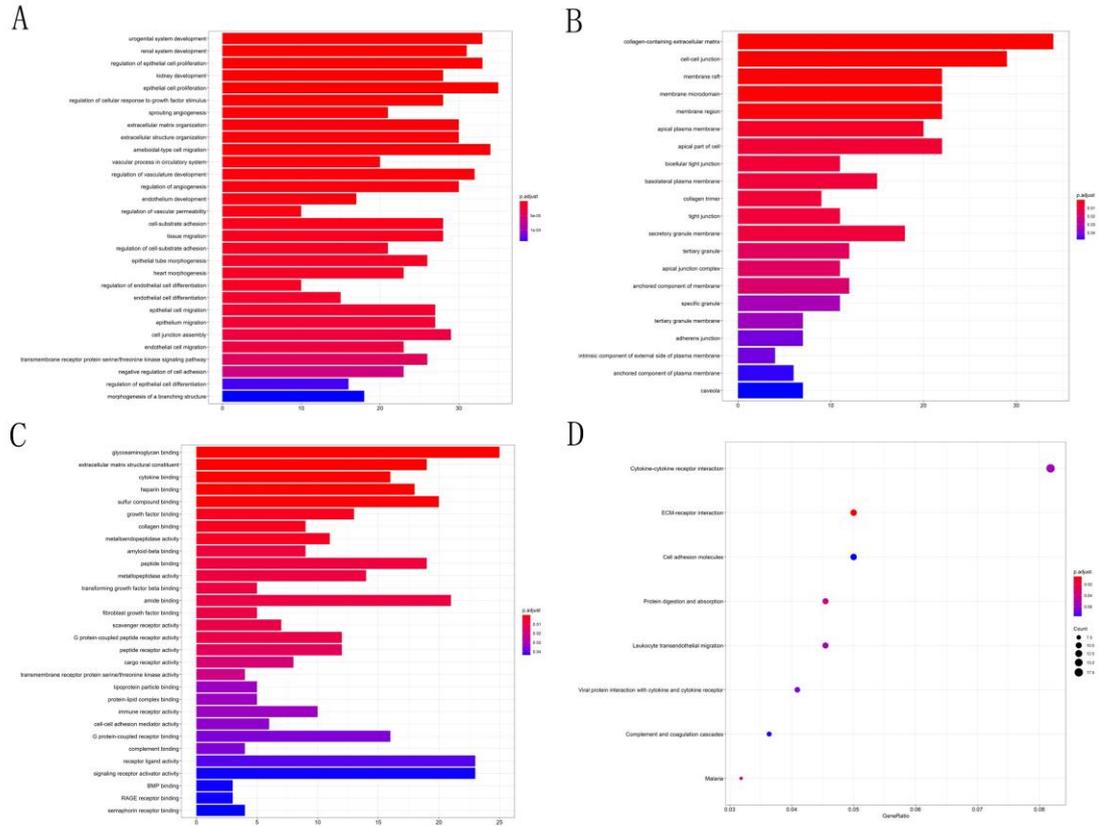


Figure 2 The GO enrichment analysis and the KEGG enrichment analysis of 500 DEGs. A: Biological process; B: Cell component; and C: Molecular function; D: Bubble plot of the KEGG enrichment analysis.

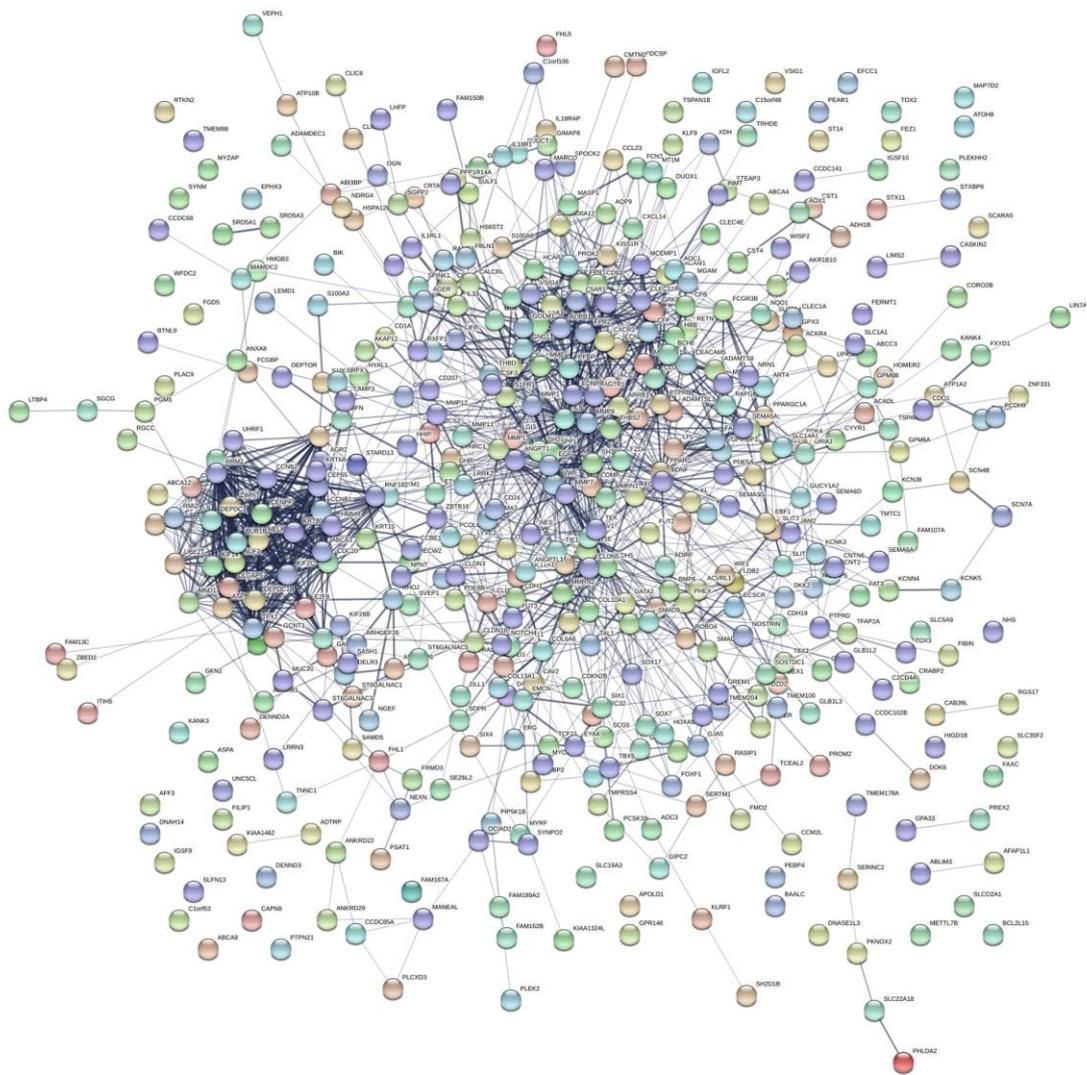


Figure 3 PPI network analysis of the DEGs using the STRING online tool.

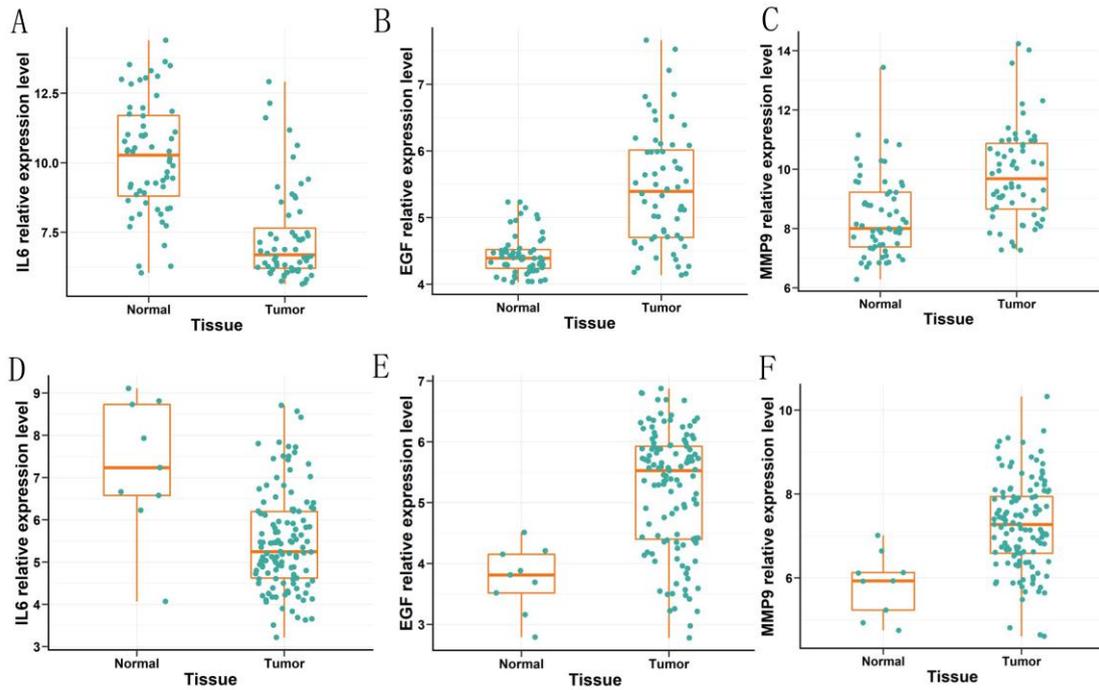


Figure 4 Analysis of the relative expression levels of IL6, EGF and MMP9 in female lung cancer samples and normal female lung tissues by bioinformatic methods. (A–F) Relative expression of IL6, EGF and MMP9 in female lung cancer tissues compared to normal tissues was analysed using GEO data sets, including series GSE19804 and GSE31210.

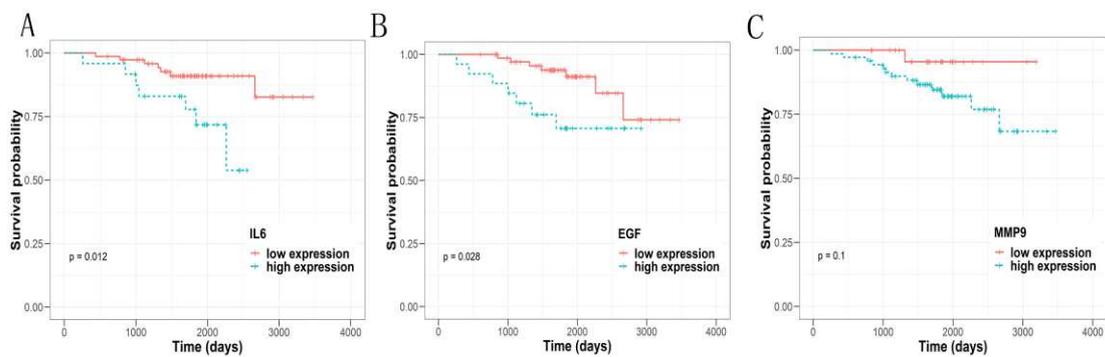


Figure 5 Kaplan-Meier survival curves for IL6, EGF and MMP9 expression in female NSCLC patients.

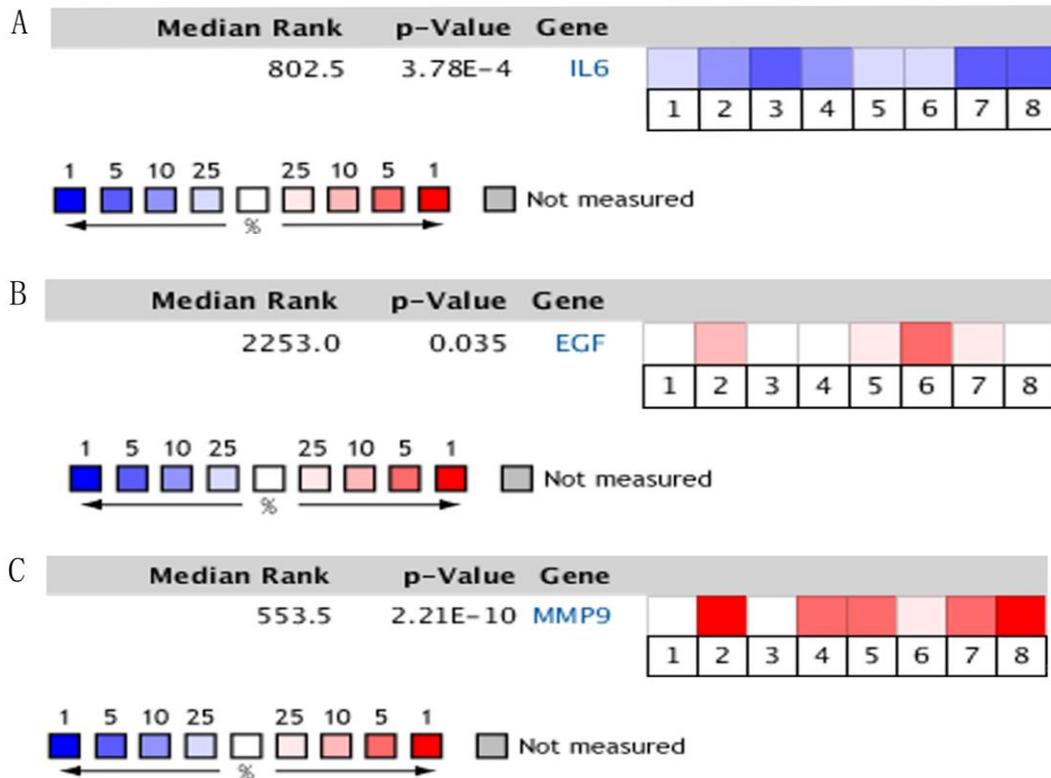


Figure 6 IL6, EGF and MMP9 expression level analysis in the Oncomine database (1 ~ 8 represent 8 different NSCLC research datasets, red represents overexpression, and blue represents low expression).

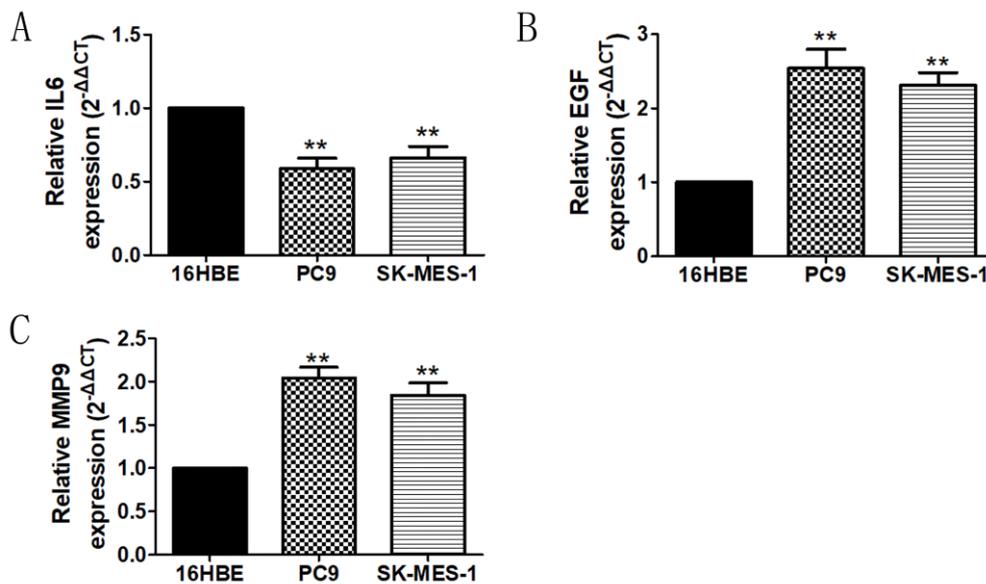


Figure 7 Analysis of the relative expression levels of IL6, EGF and MMP9 in cells (The data were obtained from at least 3 independent

experiments).

Figures

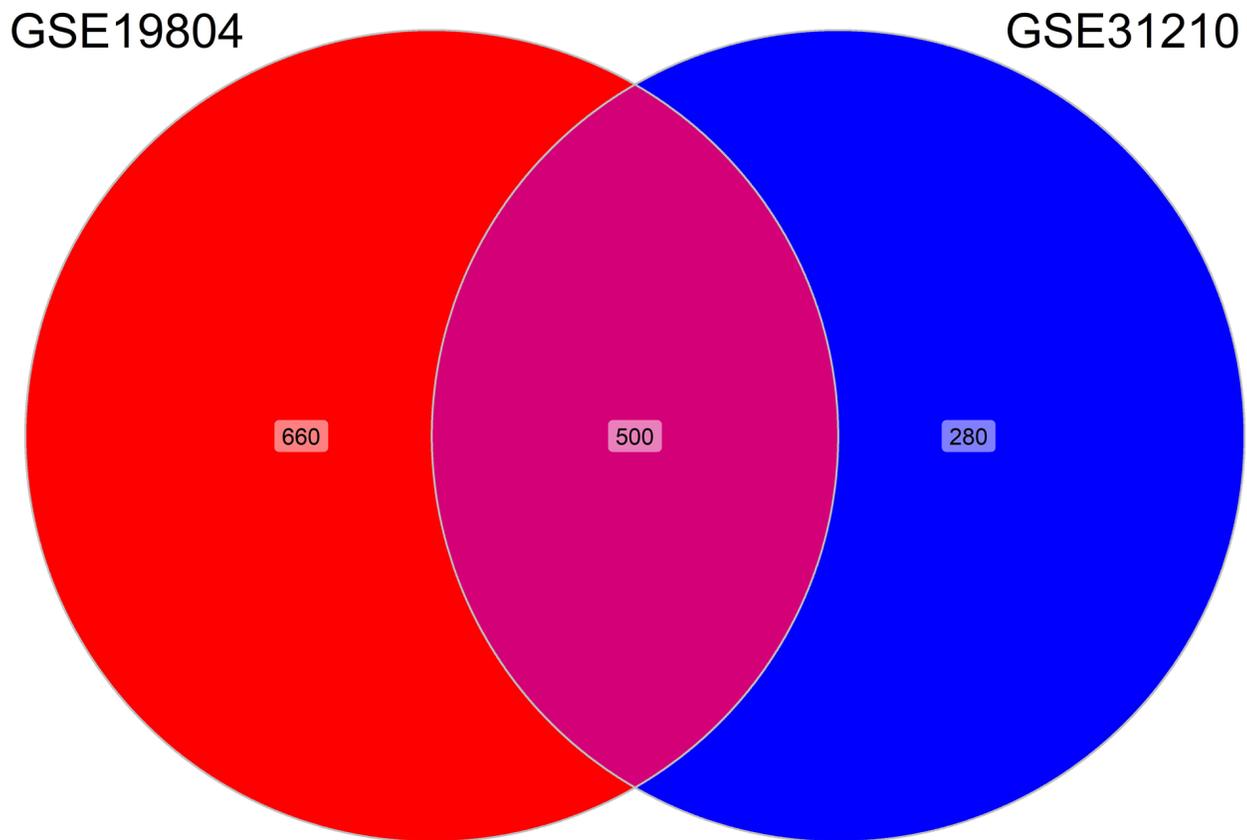


Figure 1

Five hundred common DEGs were screened from the GSE19804 and GSE31210 datasets.

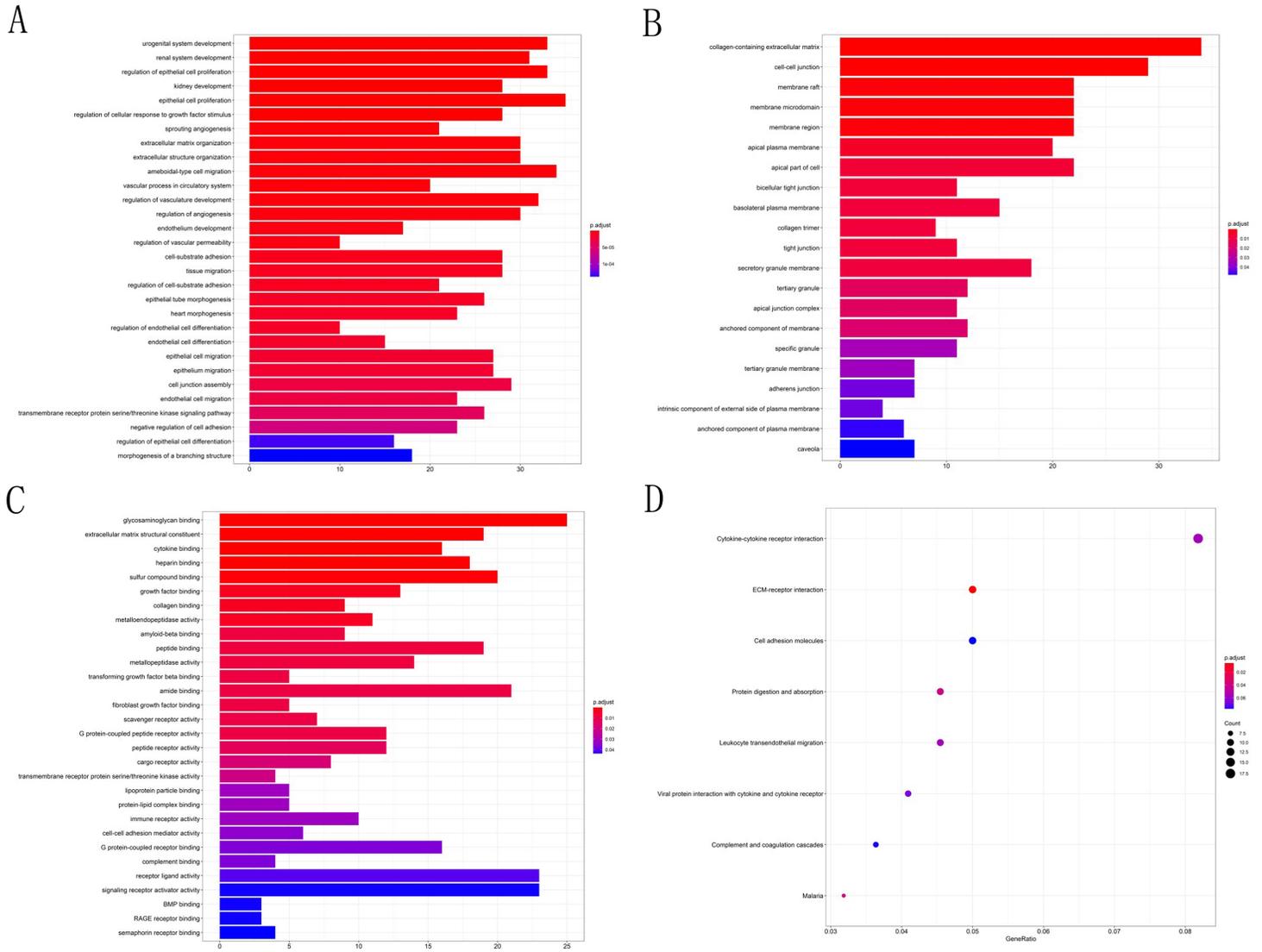


Figure 2

The GO enrichment analysis and the KEGG enrichment analysis of 500 DEGs. A: Biological process; B: Cell component; and C: Molecular function; D: Bubble plot of the KEGG enrichment analysis

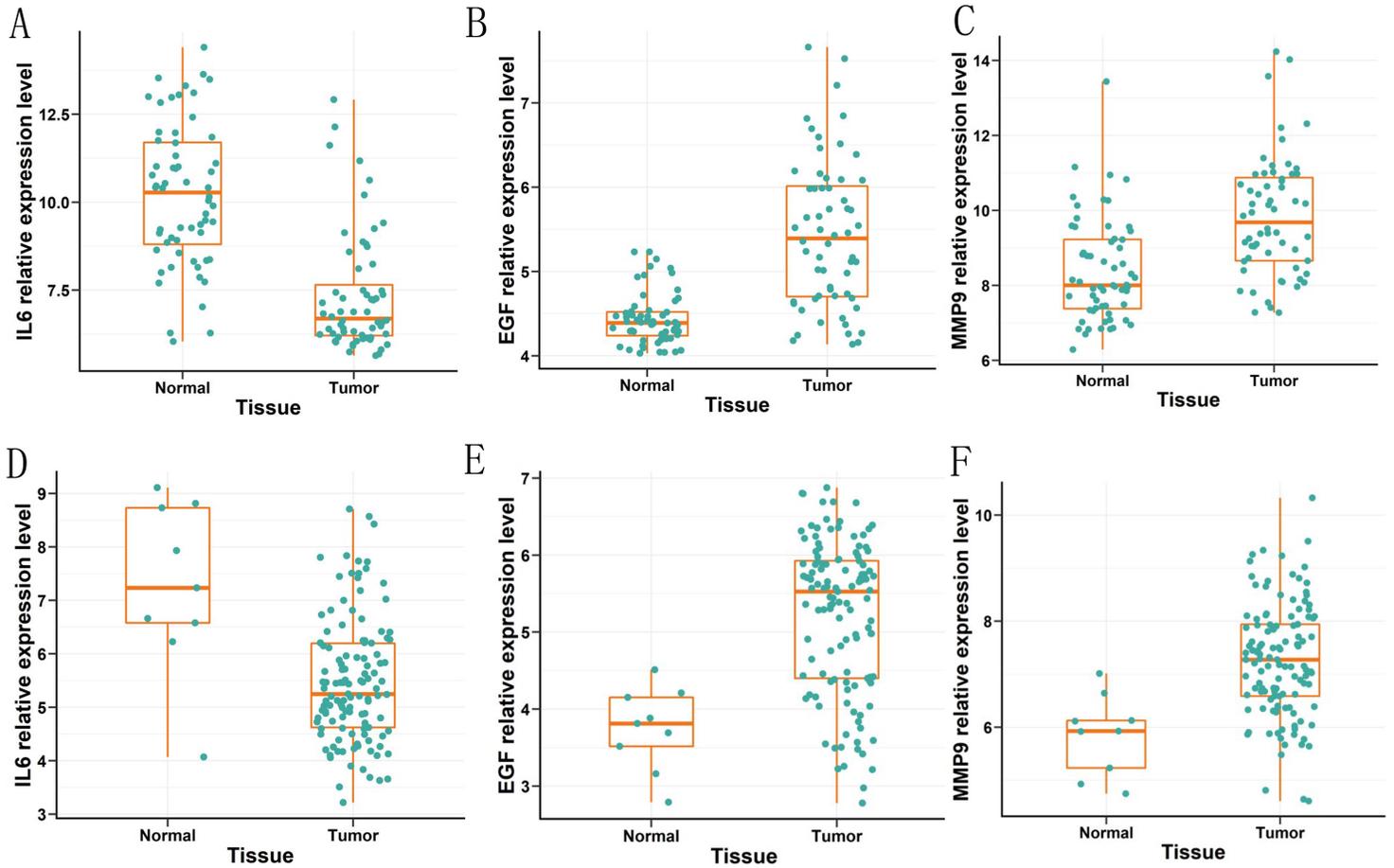


Figure 4

Analysis of the relative expression levels of IL6, EGF and MMP9 in female lung cancer samples and normal female lung tissues by bioinformatic methods. (A-F) Relative expression of IL6, EGF and MMP9 in female lung cancer tissues compared to normal tissues was analysed using GEO data sets, including series GSE19804 and GSE31210.

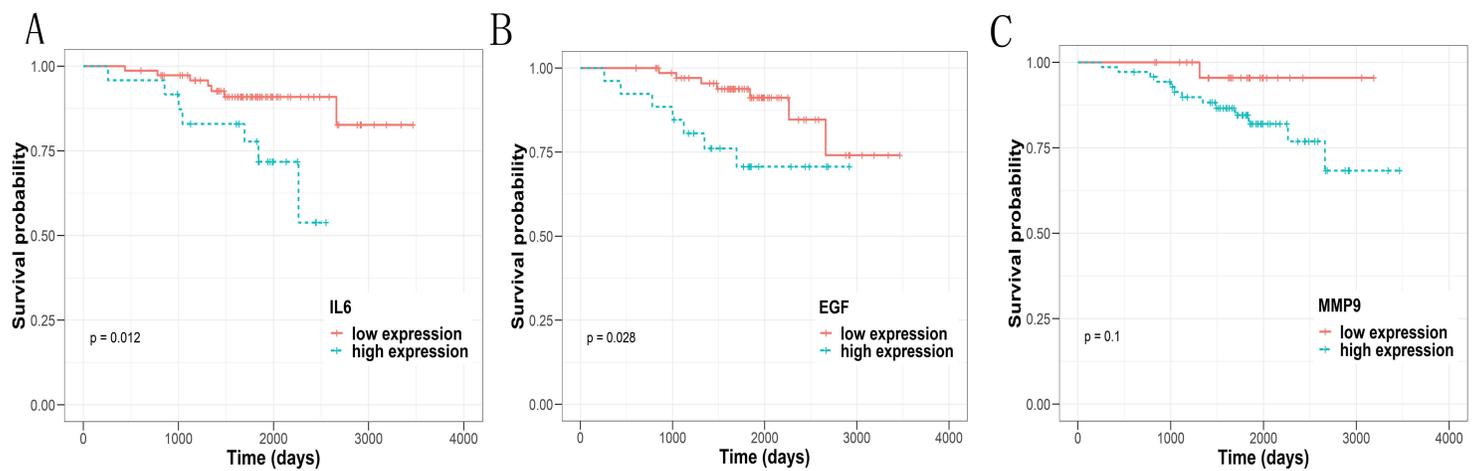


Figure 5

Kaplan Meier survival curves for IL6, EGF and MMP9 expression in female NSCLC patients.

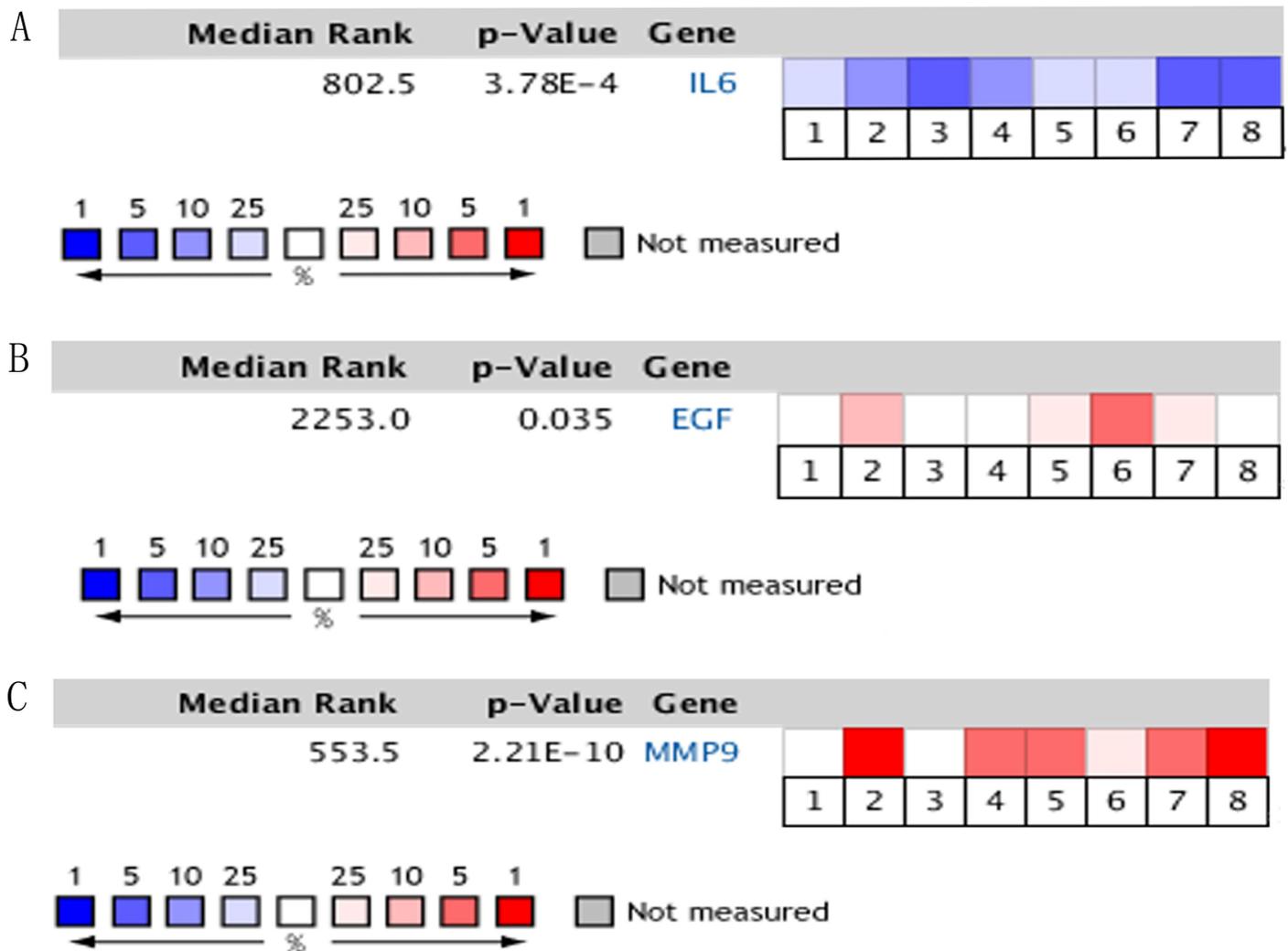


Figure 6

IL6, EGF and MMP9 expression level analysis in the Oncomine database (1-8 represent 8 different NSCLC research datasets, red represents overexpression, and blue represents low expression)

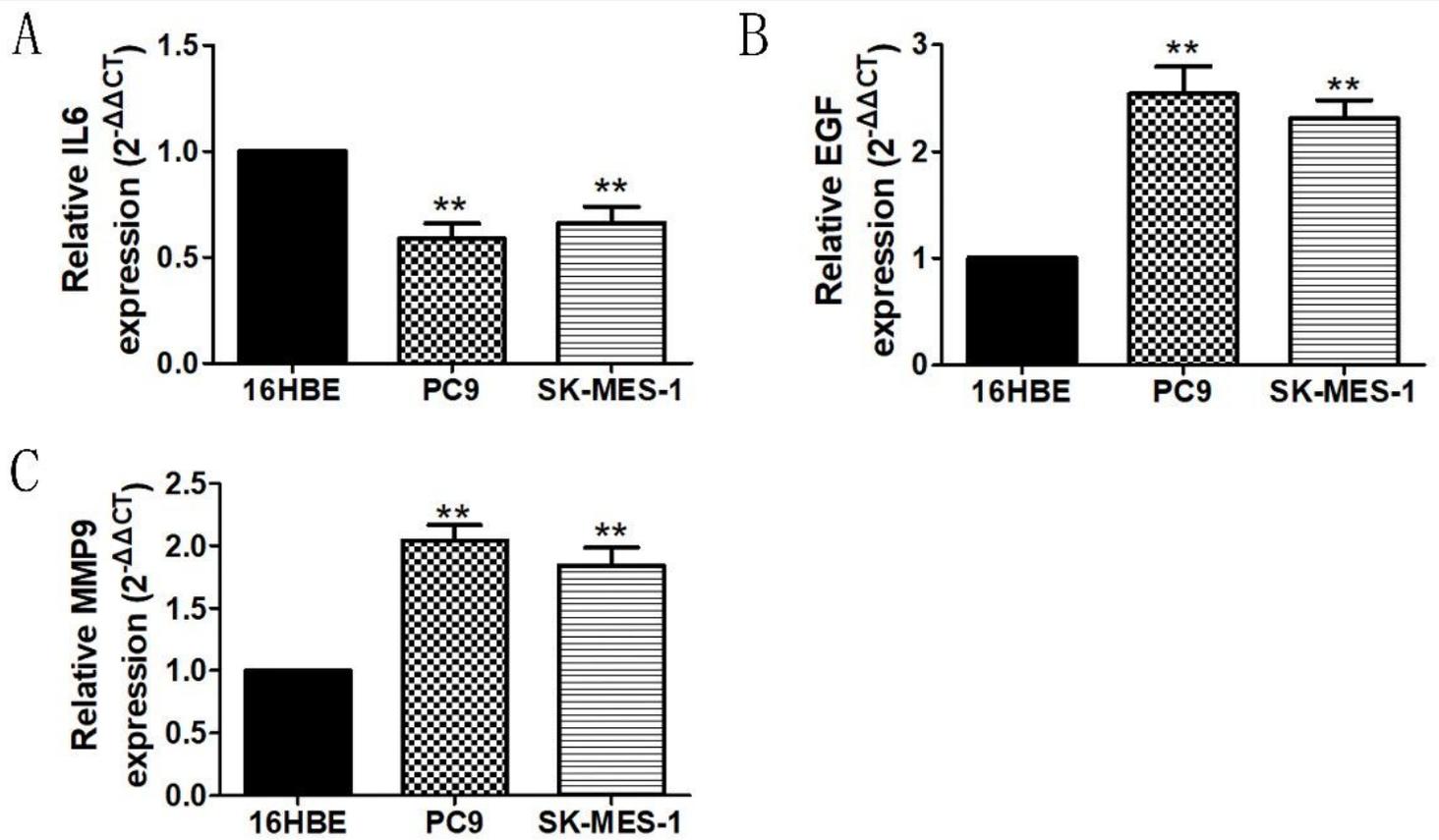


Figure 7

Analysis of the relative expression levels of IL6, EGF and MMP9 in cells (The data were obtained from at least 3 independent experiments).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [AdditionalTableS1.docx](#)
- [AdditionalTableS1.docx](#)