Int J Clin Exp Med 2015;8(8):12161-12171
www.ijcem.com /ISSN:1940-5901/IJCEM0010613

Original Article

Screening feature genes of lung carcinoma with DNA microarray analysis

Liangdong Chen, Deqiang Zhuo, Jiakuan Chen, Hongyin Yuan

Department of Oncology, Zhongnan Hospital, Wuhan University, Wuhan 430071, PR China

Received May 24, 2015; Accepted July 11, 2015; Epub August 15, 2015; Published August 30, 2015

Abstract: Lung carcinoma is the most common and aggressive malignant tumor with poor clinical outcome. Identification of new marker of lung cancer is essential for the diagnosis and prognosis of the disease. To identify differentially expressed genes (DEGs) and find associated pathways that may function as targets of lung cancer. Gene expression profiling of GSE40791 were downloaded from GEO (Gene Expression Omnibus), including 100 normal specimens and 94 lung cancer samples. The DEGs were screened out by LIMMA package in R language. Besides, novel genes associated with lung cancer were identified by co-expression analysis. Then, GO enrichment and transcription binding site analysis were performed on these DEGs, and novel genes were predicted using DAVID. Finally, PPI network was constructed by String software in order to get the hub codes involved in cancer carcinoma. A total of 541 DEGs were filtered out between normal samples and patients with lung carcinoma, including 155 up-regulated genes and 386 down-regulated genes. Additionally, nine novel genes, *CA4, CDC20, CHRDL1, DLGAP5, EMCN, GPM6A, NUSAP1, S1PR1* and *TCF21*, were figured out. The transcription biding site analysis showed that these genes were regulated by *LHX3*, *HNF3B*, *CDP*, *HFH1*, *FOXO4*, *STAT*, *SOX5*, *MEF2*, *FOXO3* and *SRY*. Hub codes as *BUB1B, MAD2L* and *TOP2A* may play as target genes in lung carcinoma in the result of PPI network analysis. Newly predicted genes and hub codes can perform as target genes for diagnose and clinical therapy of lung cancer.

Keywords: Gene expression profiling, DEG, PPI network, disease marker

Introduction

Lung carcinoma, is recognized as one of the most malignant tumor worldwide [1]. During the past 60 years, the incidence of lung carcinoma has risen sharply in all developed countries all over the world and in fact, it's one of the most common causes of death among patients with cancer. Worse still, as for the prognosis of patients with lung cancer, X-ray and CT methods are far from enough in clinical mediation. Since there hasn't been much research about the molecular mechanism of the disease, prognostic and treatment of lung cancer become much harder [2-4]. Therefore, research of new target genes in lung carcinoma can be quite important and beneficial for the improving of the clinical management of lung cancer.

The development of microarray technology has provided new insights into cancer diagnosis and treatment [5]. It's been reported that many biomarkers associated with lung cancer were selected out based on the data of gene expression profiles. Combining with bio-networks (PPI network; signal network; regulatory network; metabolic network), the complex pathogenesis of lung cancer are analyzed [6, 7]. It has been proved that this simple technique can be useful in the early and accurate diagnosis of lung cancer [8]. What's more, the selected differentially expressed genes (DEGs) can perform as keys to understanding what goes wrong in some certain conditions (cancer, stress, etc). In other cases, these genes can be used as "features" for classifier [9]. In many bioinformatics study, screening DEGs is usually served as the starting point of the whole research [10, 11].

Therefore, to better understand the mechanism of lung cancer, we aim to find cancer-related genes and pathways by analyzing microarray data from GEO database. Besides, we constructed protein-protein interaction networks to investigate the critical DEGs in the progression. By detecting hub nodes of the network, the

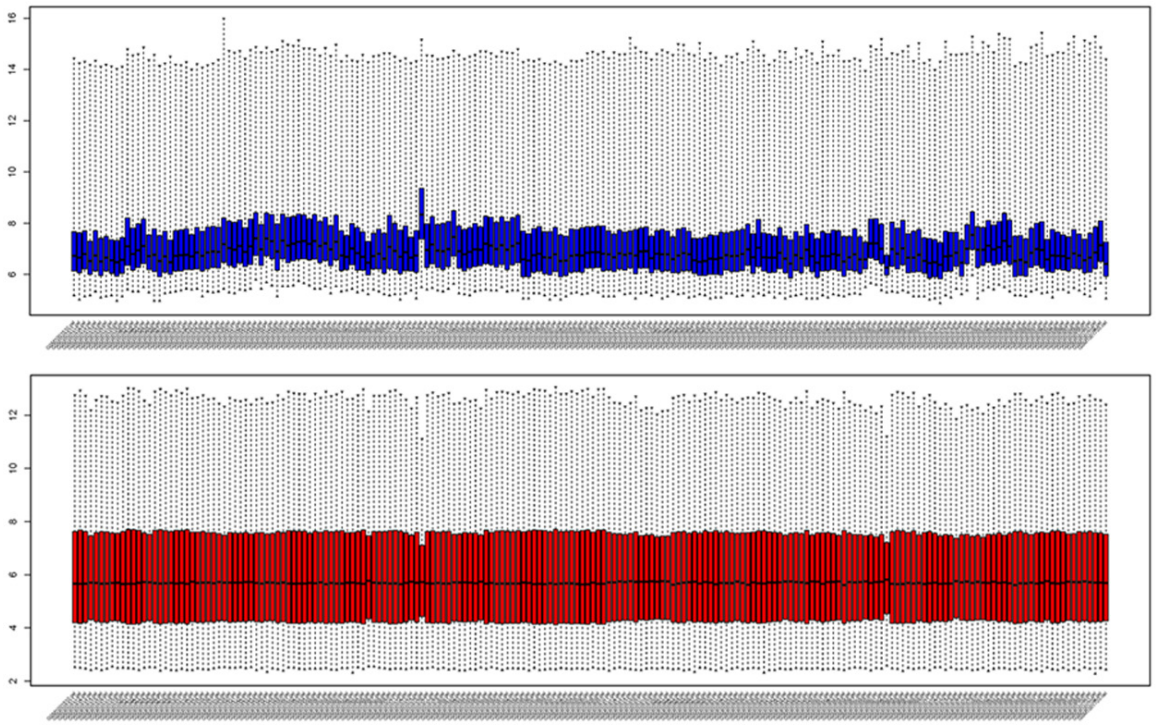

Figure 1. Box plots for expression data before normalization (left) and after normalization (right). The horizontal axis is the name of samples while the vertical axis stands for the values of expression. The black line in the cassette is the median of data in each group, which represents the degree of normalization. The black line in the right figure was almost on the same line, indicating an excellent degree of normalization.

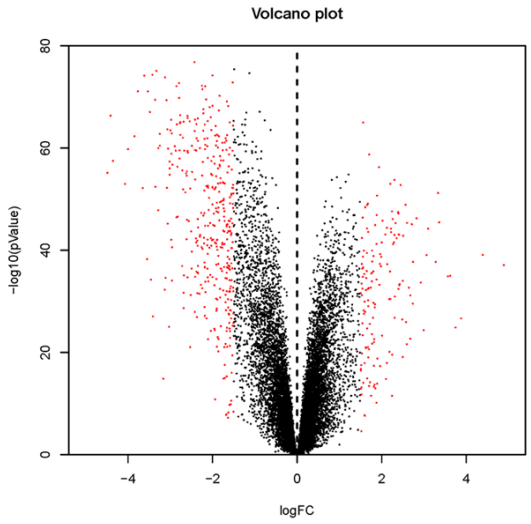

Figure 2. The volcano plot of differentially expressed genes. The abscissa is logFC and the ordinates is -log10 (*P* Value). The red dots stand for the differentially expressed genes while the black dots represent genes not differentially expressed.

most essential targets of lung cancer were further analyzed for the lung cancer medication in clinical therapy.

Material and methods

Data source

The microarray profiles of lung cancer were extracted from GEO (Gene Expression Omnibus, http://www.ncbi.nlm.nih.gov/geo/) database under the accession number of GSE40791 [12]. A total of 194 specimens, including 100 gene chips from normal tissues and 94 gene chips from tissues of lung carcinoma patients, were available for the analysis.

Data preprocessing

The original data were preprocessed by Affymetrix package [13] in R language. The original CEL files were converted into probe expression measures. The differentially expressed genes between the expression profiles of 94 lung carcinoma patients and 100 normal samples were identified by LIMMA package [14], with the 100 normal samples as control group.

The probe-level data was converted into gene names based on the GPL570 platform. Probes

Table 1. The differentially expressed genes (DEGs)

Gene Name	ENTREZ_GENE_ID	Normal Ave	Tumor Ave	logFC	t	*p* Value	q Value
AGER	177	11.25428	6.788105	-4.46618	-37.3906	8.11E-91	1.67E-86
CA4	762	9.41108	5.973918	-3.43716	-34.5028	6.60E-85	6.81E-81
GPM6A	2823	8.045994	4.891125	-3.15487	-33.9418	1.02E-83	7.03E-80
RTKN2	219790	6.533756	4.106482	-2.42727	-31.1295	1.55E-77	8.02E-74
FAM107A	11170	9.847733	6.521242	-3.32649	-30.3736	8.27E-76	2.85E-72
C13orf36	400120	6.947827	3.527286	-3.42054	-30.0458	4.73E-75	1.22E-71
ACVRL1	94	8.812331	6.811757	-2.00057	-29.9841	6.58E-75	1.50E-71
NCKAP5	344148	7.609707	3.995841	-3.61387	-29.9656	7.26E-75	1.50E-71
TCF21	6943	9.596483	6.474893	-3.12159	-29.8241	1.55E-74	2.91E-71
ADAMTSL3	57188	7.233111	5.713201	-1.51991	-29.3969	1.55E-73	2.66E-70

Table 2. Cancer-related genes

Term	*P* Value	Genes
CANCER	0.002242	CYP24A1, PGC, MMP9, SLC6A4, PPARG, EZH2, MMP7, TTK, ADH1B, SPINK1, CXCR2, MMRN1, CXCL12, MMP1, GPX2, AGTR1, EDNRB, AGTR2, NQO1, TOP2A, GHR, MMP13, SLIT2, MMP12, ADRB2, MAD2L1, CD36, NME1, SSTR1, BUB1B, CAV2, CAV1, CHEK1, CX3CL1, ZBTB16, SFTPA1, TIMP3, MDK, ACE, FMO2, FMO3, SPP1, LPL, CES1, BIRC5, AGER, ABCG2, KDR, CXCL14, PECAM1, TGFBR3, ID3, SELE

Table 3. Co-expression relationships

node1	node2	adjacency cor
ACVRL1	CDH5	0.754042
ACVRL1	RAMP2	0.785929
ACVRL1	ROBO4	0.801125
ADH1B	CHRDL1	0.767314
AGER	CA4	0.770725
AGER	GPM6A	0.82897
BIRC5	CDC20	0.75286
BIRC5	DLGAP5	0.797217
BUB1B	DLGAP5	0.764016
BUB1B	NUSAP1	0.764438

matching more than one gene were eliminated and average value was used for probes matching the same one gene.

Screening of differentially expressed genes (DEGs)

Limma package of R language [14] was used to identify the DEGs between lung cancer samples and normal specimens. Fold change > 1.5 was used as the threshold to determine the significance of gene expression difference. FDR (False Discovery Rate) of q-value was adjusted to 0.05. Heatmaps were made to ensure that the screened genes have significant differences.

Prediction of novel genes related with lung carcinoma

Differentially expressed genes play a role through interacting with each other. In this study, we used DAVID (Database for Annotation Visualization and Integrated Discovery) database [15] to perform the GENTIC analysis to search the lung-carcinoma related genes. *P*-Value < 0.05 was used as the threshold.

WGCNA (Weighted Correlation Network Analysis) [16] was used to construct the co-expression network among the differentially expressed genes (adjacency threshold is 0.75). Cytoscape [17] was used to visualize the co-expression interaction relationships.

GO enrichment analysis

Gene ontology (GO) has become a commonly used approach for functional studies of large-scale genomic or transcriptomic data. GO enrichment analysis was performed by DAVID [15], a high-throughput and integrated data-mining environment, for the lung carcinoma related genes and predicted ones. FDR=0.05 was used as the cut-off criterion.

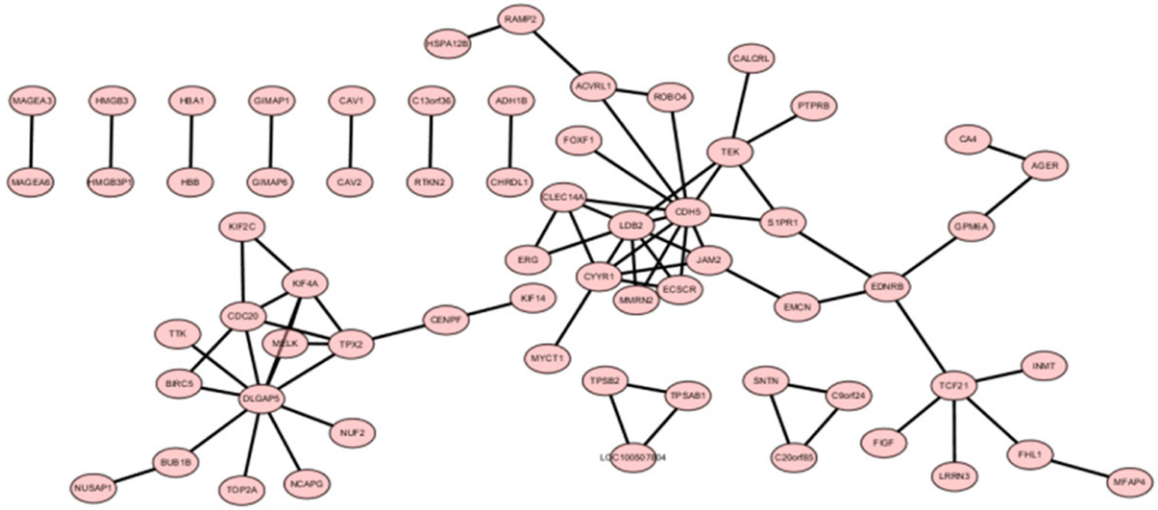

Figure 3. Results of co-expression analysis.

Transcription factor binding sites analysis

UCSC_TFBS analysis was performed on all the cancer-related genes by DAVID software to predict the transcription factors regulating these DEGs. *P*-Value < 0.05 was used as the threshold.

Protein-protein interaction network (PPI) analysis

String (http://string-db.org/) [18], a database of known and predicted protein-protein interactions, is accessible through the Internet. Information of human PPI relationships and genes related to some diseases in String were more abundant than other PPI databases. Therefore, the PPI from String were collected for the construction of differential protein interaction network among the cancer related genes and newly predicted genes in lung cancer. The DEGs were mapped to the String database and then known and predicted associations were scored and integrated. Hereinto, combined-score > 0.4 was the threshold. What's more, the DEGs were visualized by Cytoscape software. Finally, interaction network was constructed by integrating these relationships.

Results

Data preprocessing and DEGs screening

Systematic bias among original data was removed after preprocessing by Affymetrix package in R language. Well normalized gene expression data were obtained (**Figure 1**). The black line in each box was approximately at the same level, indicating an excellent degree of standardization. After preprocessing, the normalized expression profile data were differentially compared and a total of 541 differentially expressed genes exceeding the difference threshold (q value < 0.05 and FC > 1) were screened out, including 155 up-regulated genes and 386 down-regulated genes (**Figure 2**). The top ten DEGs were listed in **Table 1**.

Prediction of novel cancer-related genes

A total of 53 cancer-related genes were figured out, which is listed in **Table 2**. Nine genes, CA4, CDC20, CHRDL1, DLGAP5, EMCN, GPM6A, NUSAP1, S1PR1 and TCF21, were identified as differentially expressed genes in our study, which, meanwhile may serve as genes closely related with lung carcinoma. *CA4*, *CDC20*, *CHRDL1*, *DLGAP5*, *GPM6A*, *NUSAP1* were selected as the novel cancer-related genes by the result of co-expression table constructed by WGCNA method (**Table 3**). Besides, these genes were mapped to co-expression figure by Cytoscape software (**Figure 3**).

Functional analysis of the DEGs in lung cancer

GO enrichment analysis was performed on the DEGs and novel genes associated with lung carcinoma (**Figure 4**, **Table 4**). The most

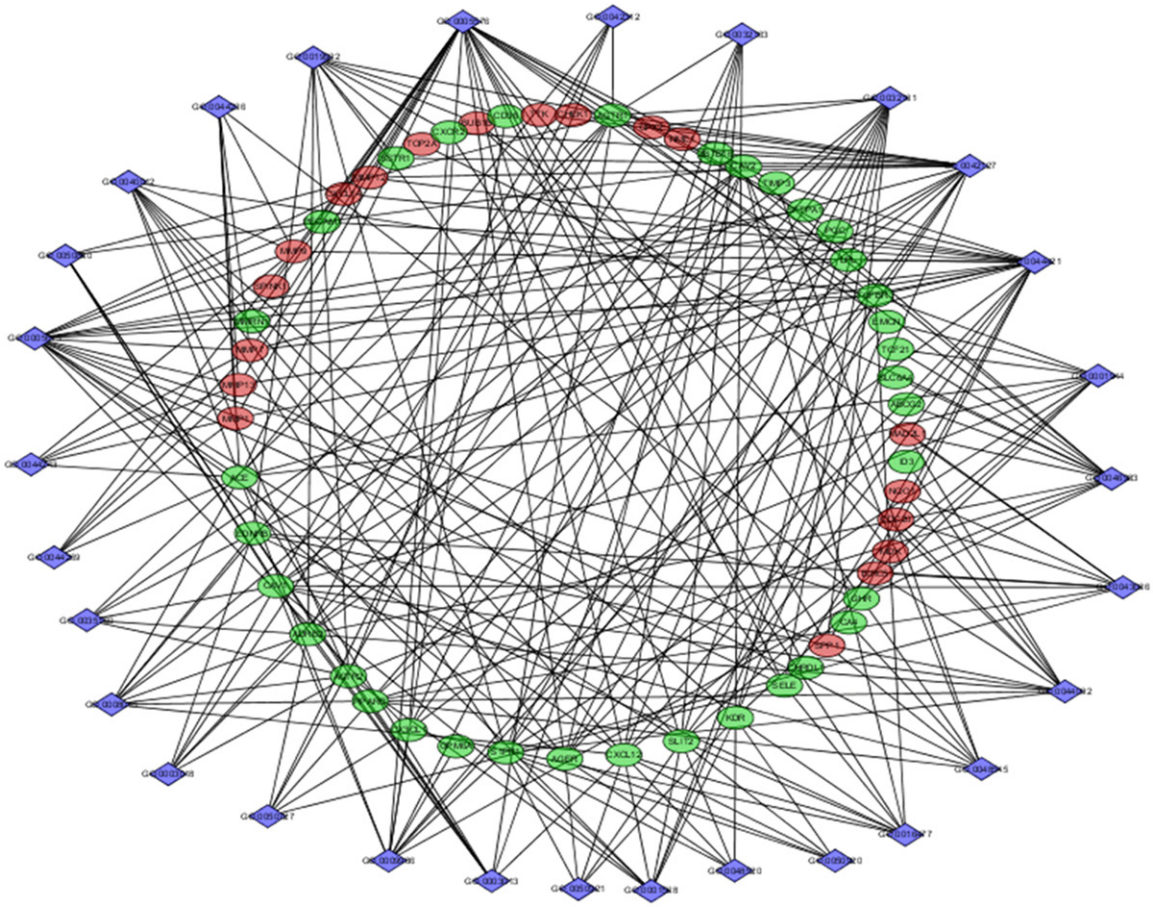

Figure 4. GO enrichment analysis.

Table 4. New cancer-related feature genes

Term	New Genes
cancer	CA4, CDC20, CHRDL1, DLGAP5, EMCN, GPM6A, NUSAP1, S1PR1, TCF21

Table 5. GO Enrichment analysis

Term	FDR
GO:0032101~regulation of response to external stimulus	5.19E-09
GO:0044421~extracellular region part	1.82E-05
GO:0042127~regulation of cell proliferation	3.79E-05
GO:0005615~extracellular space	3.49E-05
GO:0019932~second-messenger-mediated signaling	1.01E-04
GO:0032103~positive regulation of response to external stimulus	5.34E-04
GO:0044236~multicellular metabolic process	8.49E-04
GO:0001944~vasculature development	0.002062975
GO:0008015~blood circulation	0.002185315
GO:0003013~circulatory system process	0.002185315

enriched GO terms among the cancer-related DEGs were obtained (**Table 5**). We found that they were mainly involved in the regulation of cell proliferation and external stimulus.

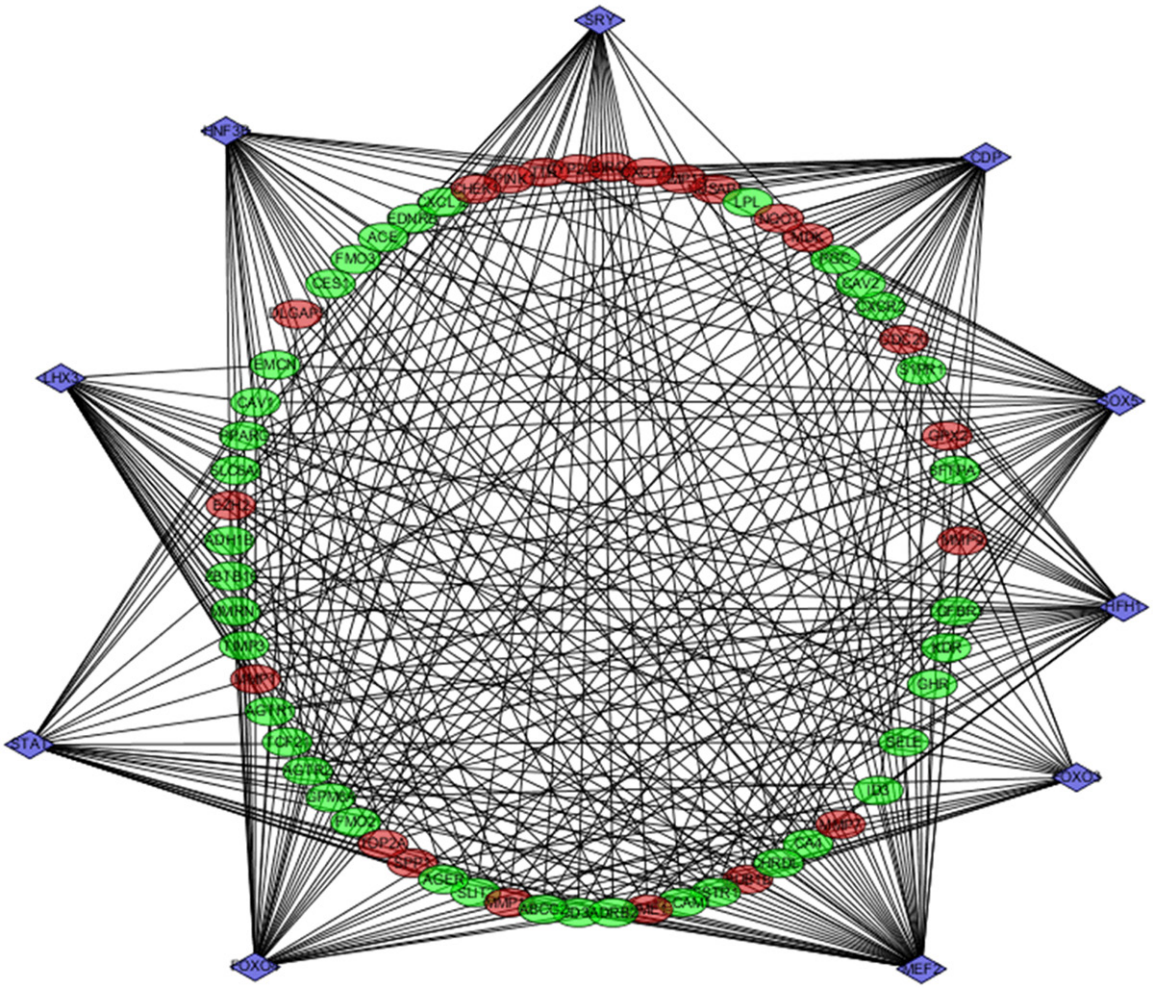

Figure 5. Analysis of binding sites of transcription factors. The rhombus is binding sites of transcription factors and the circle represents genes. The red circle is up-regulated genes while the green ones represent down-regulated genes.

Transcription factor binding site analysis

LHX3, HNF3B, CDP, HFH1, FOXO4, STAT, SOX5, MEF2, FOXO3, SRY were found to mainly take part in the regulation of these DEGs by the transcription binding sites analysis (**Figure 5**, **Table 6**).

Protein-protein interaction network analysis

PPI network was constructed by Cytoscape software based on the PPIs between DEGs and genes or proteins (**Figure 6**). Hub nodes as *BUB1B*, *MAD2L* and *TOP2A*, were identified as the most essential genes in lung cancer disease (**Figure 7**), which may also function as the new targets for medicine treatment.

Discussion

Key genes and pathways associated with lung carcinoma were disclosed by using bioinformatics methods. By differentially expressed genes (DEGs) analysis and database for annotation, visualization and integrated discovery (DAVID), we discovered novel markers and potential targets of lung cancer, which will be very helpful for the prognosis and therapy of lung carcinoma. Nine novel genes, including *CA4*, *CDC20*, and *TCF21*, which were also identified as DEGs may serve as feature genes closely related with lung carcinoma.

At the top of the ranked listed differentially expressed genes among the patients with lung cancer, *CA4* and *TCF21* happen to be the genes

Table 6. Analysis of transcription binding sites

Term	*P* Value	Genes
LHX3	0.004091	EMCN, CAV1, PPARG, SLC6A4, EZH2, ADH1B, ZBTB16, MMRN1, TIMP3, MMP1, AGTR1, TCF21, AGTR2, GPM6A, FMO2, TOP2A, GHR, SPP1, AGER, SLIT2, MMP12, ABCG2, ADRB2, CD36, CHRDL1, SSTR1, BUB1B, TGFBR3, ID3, SELE
HNF3B	0.010339	CYP24A1, CAV1, PPARG, TTK, SPINK1, CHEK1, ZBTB16, TIMP3, CXCL12, EDNRB, TCF21, AGTR2, ACE, GPM6A, FMO3, TOP2A, GHR, CES1, DLGAP5, BIRC5, AGER, SLIT2, MMP12, ADRB2, CHRDL1, NME1, SSTR1, PECAM1, CA4, TGFBR3, SELE
CDP	0.01474	CAV2, CYP24A1, EMCN, CAV1, PGC, PPARG, SLC6A4, EZH2, TTK, ADH1B, SPINK1, CHEK1, CXCR2, ZBTB16, MMRN1, MDK, TIMP3, CXCL12, EDNRB, ACE, AGTR2, GPM6A, NQO1, TOP2A, GHR, LPL, NUSAP1, MMP13, AGER, SLIT2, MMP12, ABCG2, CD36, CHRDL1, CXCL14, NME1, SSTR1, BUB1B, TGFBR3, CA4, ID3
HFH1	0.023739	CAV2, CYP24A1, MMP9, EZH2, PPARG, ZBTB16, SFTPA1, MMRN1, TIMP3, MDK, CXCL12, MMP1, GPX2, AGTR1, EDNRB, TCF21, AGTR2, ACE, GPM6A, FMO3, NQO1, TOP2A, GHR, MMP13, SLIT2, ADRB2, CD36, CHRDL1, CA4, TGFBR3, ID3, SELE
FOXO4	0.024888	CAV2, CYP24A1, CAV1, MMP9, PGC, PPARG, SLC6A4, EZH2, SPINK1, CHEK1, ZBTB16, MMRN1, MDK, TIMP3, CXCL12, AGTR1, EDNRB, AGTR2, ACE, GPM6A, NQO1, GHR, CES1, DLGAP5, NUSAP1, MMP13, AGER, SLIT2, MMP12, ABCG2, CHRDL1, CXCL14, SSTR1, PECAM1, TGFBR3, CA4, ID3
STAT	0.032417	CAV2, EMCN, CAV1, EZH2, PPARG, MMP7, CHEK1, ZBTB16, SFTPA1, TIMP3, EDNRB, AGTR2, GPM6A, TOP2A, SPP1, AGER, MMP12, SLIT2, ADRB2, CD36, CHRDL1, SSTR1, CA4, TGFBR3, ID3, SELE
SOX5	0.03526	CAV2, EMCN, CAV1, EZH2, PPARG, TTK, CHEK1, ZBTB16, MMRN1, MDK, TIMP3, AGTR1, EDNRB, AGTR2, S1PR1, GPM6A, FMO2, FMO3, NQO1, TOP2A, GHR, CDC20, MMP13, SLIT2, MMP12, ABCG2, ADRB2, CHRDL1, NME1, SSTR1, PECAM1, BUB1B, TGFBR3
MEF2	0.038847	CAV2, CYP24A1, EMCN, CAV1, MMP9, PGC, SLC6A4, PPARG, EZH2, TTK, CHEK1, ZBTB16, MMRN1, MDK, TIMP3, CXCL12, MMP1, GPX2, TCF21, AGTR1, EDNRB, ACE, AGTR2, S1PR1, GPM6A, FMO3, NQO1, TOP2A, SPP1, GHR, LPL, CES1, DLGAP5, NUSAP1, CDC20, AGER, SLIT2, MMP12, ADRB2, CD36, CHRDL1, CXCL14, NME1, BUB1B, TGFBR3, ID3, SELE
FOXO3	0.045322	EMCN, EZH2, PPARG, ZBTB16, MDK, TIMP3, SLIT2, MMP12, KDR, GPX2, EDNRB, ADRB2, S1PR1, CHRDL1, GPM6A, NME1, PECAM1, CA4, TGFBR3, SPP1, GHR
SRY	0.04934	CYP24A1, CAV2, EMCN, CAV1, PPARG, EZH2, ADH1B, TTK, CHEK1, ZBTB16, MMRN1, TIMP3, AGTR1, EDNRB, AGTR2, GPM6A, TOP2A, GHR, SLIT2, MMP12, ABCG2, ADRB2, CD36, CHRDL1, SSTR1, NME1, PECAM1, BUB1B, TGFBR3

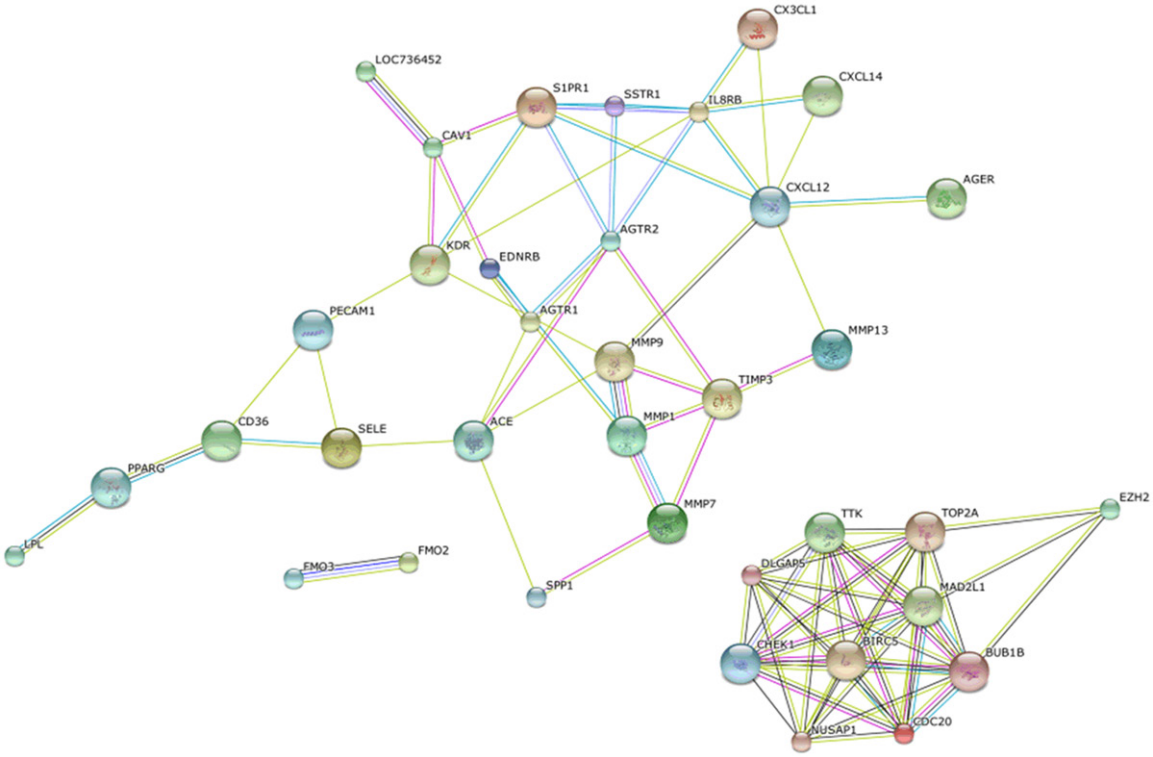

Figure 6. PPI network of new predicted cancer-related genes. The circle stands for genes and the line indicates the interactions among genes. The interior of the circle represents the structure of proteins. The color of the line provides evidence of the different interactions among proteins. (A red line indicates the presence of fusion evidence; a green line, neighborhood evidence; a blue line, concurrence evidence; a purple line, experimental evidence; a yellow line, text mining evidence; a light blue line, database evidence; a black line, coexpression evidence).

which are most likely to play as the new targets of cancer-related genes. TCF 21 (transcription factor 21) encodes a protein expressed in mesenchyme-derived tissues of lung, gut and kidney [19]. It has been reported that the disease of clear cell renal cell carcinoma is associated with TCF21, providing evidence of its role as potential target in lung cancer disease [20]. The GO enrichment analysis revealed that DEGs were significantly enriched in response to external stimulus, the function of which was mostly performed by relating proteins on plasma membrane. Given that TCF21 is involved in the epithelial differentiation and morphogenesis in kidney and lung [21], its potential role in the occurrence of lung cancer are worthy of future research. What's more, according to the results of our analysis in the binding sites of transcription factors, we found that the gene TCF21 was mainly regulated by LHX3 and HNF3B. Thus, the imported role of TCF21 in pathogenesis of lung cancer was confirmed by our bioinformatics methods.

Besides, PPI network was also constructed to visualize the interactions between DEGs and other genes. Three hub codes, *BUB1B*, *MAD2L1* and *TOP2A*, were found to have more interactions with other DEGs, suggesting their potential role as the most essential genes in lung carcinoma-related genes. *BUB1B* (BUB1 mitotic checkpoint serine/threonine kinase B) is involved in spindle checkpoint function and protein kinase activity [22]. The binding of *CDC20*, the new-found feature gene related with lung cancer, to APC/C (anaphase-promoting complex/cyclosome) can be blocked by this gene, thus delaying the onset of anaphase and ensuring proper chromosome segregation [23, 24]. It has been found that impaired spindle checkpoint function is associated with many forms of cancer [25]. Therefore, we predict that *BUB1B* may serve as the new target in the diagnosis of lung carcinoma. *MAD2L1* (MAD2 mitotic arrest deficient-like 1) is a component of the mitotic spindle assembly checkpoint, so we predict that its role in lung carcinoma is similar

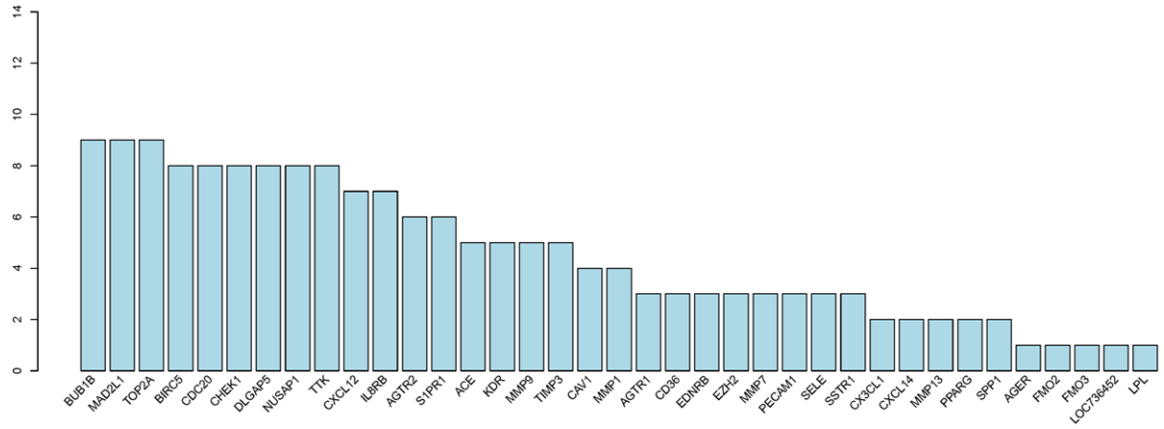

Figure 7. Histogram of numbers of genes adjacent to interaction network.

to that of gene BUB1B [26], which, of course, need further research. What's more, their function in cell proliferation is consistent with the GO enrichment analysis, that is, the significantly enriched GO term 0042127: the regulation of cell proliferation. TOP2A encodes a DNA topoisomerase that's involved in DNA transcription and replication. Topoisomerase IIalpha (TopoII alpha), one type of topoisomerase, was reported to have a relationship with lung adenocarcinoma [27]. And patients survived from lung adenocarcinooma were found to be with low content of TopoII alpha. Thus, the content of TOP2A in patients with lung carcinoma has prognostic value. Earlier reports about their roles in cancer hadn't been made clear [28-30]. Our research further confirmed their potential role as therapy target in clinical mediation.

In conclusion, we identified nine novel genes associated with lung carcinoma. Some of them may be important players in pathogenesis of lung cancer. What's more, hub genes as BUB1B, MAD2L1 and TOP2A, which may provide new views on diagnosis and prognosis of patients with lung cancer, were selected out, indicating their potential use as medication targets in clinical outcome of cancer disease. However, there is still a limitation in our present study, which means there is still need to conduct further research about lung carcinoma for disclosing the molecular mechanism and advancing the therapy development.

Acknowledgements

This study is supported by the Natural Science Foundation of China (81000974) and the Ph.D. Programs Foundation of Ministry of Education of China (20100141120015).

Disclosure of conflict of interest

None.

Address correspondence to: Dr. Liangdong Chen, Department of Oncology, Zhongnan Hospital, Wuhan University, 169 Donghu Road, Wuchang District, Wuhan 430071, PR China. Tel: 86-02767813151; Fax: 86-02767813151; E-mail: Zqcld@163.com

References

[1] Breuer R, Postmus P and Smit E. Molecular pathology of non-small-cell lung cancer. Respiration 2004; 72: 313-330.

[2] Esteller M, Sanchez-Cespedes M, Rosell R, Sidransky D, Baylin SB and Herman JG. Detection of aberrant promoter hypermethylation of tumor suppressor genes in serum DNA from non-small cell lung cancer patients. Cancer Res 1999; 59: 67-70.

[3] Eberhard DA, Johnson BE, Amler LC, Goddard AD, Heldens SL, Herbst RS, Ince WL, Jänne PA, Januario T and Johnson DH. Mutations in the epidermal growth factor receptor and in KRAS are predictive and prognostic indicators in patients with non-small-cell lung cancer treated with chemotherapy alone and in combination with erlotinib. J Clin Oncol 2005; 23: 5900-5909.

[4] Paez JG, Jänne PA, Lee JC, Tracy S, Greulich H, Gabriel S, Herman P, Kaye FJ, Lindeman N and Boggon TJ. EGFR mutations in lung cancer: correlation with clinical response to gefitinib therapy. Science 2004; 304: 1497-1500.

[5] Pastorino U. Lung cancer screening. Br J Cancer 2010; 102: 1681-1686.

[6] Ouyang Y, Huang X, Lu Z and Yao J. Genomic survey, expression profile and co-expression network analysis of OsWD40 family in rice. BMC Genomics 2012; 13: 100.
[7] Fu J, Tang W, Du P, Wang G, Chen W, Li J, Zhu Y, Gao J and Cui L. Identifying microRNA-mRNA regulatory network in colorectal cancer by a combination of expression profile and bioinformatics analysis. BMC Syst Biol 2012; 6: 68.
[8] Gordon GJ, Jensen RV, Hsiao L-L, Gullans SR, Blumenstock JE, Ramaswamy S, Richards WG, Sugarbaker DJ and Bueno R. Translation of microarray data into clinically relevant cancer diagnostic tests using gene expression ratios in lung cancer and mesothelioma. Cancer Res 2002; 62: 4963-4967.
[9] Franklin WA. Diagnosis of lung cancer: pathology of invasive and preinvasive neoplasia. Chest 2000; 117: 80S-89S.
[10] Reiner A, Yekutieli D and Benjamini Y. Identifying differentially expressed genes using false discovery rate controlling procedures. Bioinformatics 2003; 19: 368-375.
[11] Pan W. A comparative review of statistical methods for discovering differentially expressed genes in replicated microarray experiments. Bioinformatics 2002; 18: 546-554.
[12] Zhang Y, Foreman O, Wigle DA, Kosari F, Vasmatzis G, Salisbury JL, van Deursen J and Galardy PJ. USP44 regulates centrosome positioning to prevent aneuploidy and suppress tumorigenesis. J Clin Invest 2012; 122: 4362-4374.
[13] Gautier L, Cope L, Bolstad BM and Irizarry RA. affy-analysis of Affymetrix GeneChip data at the probe level. Bioinformatics 2004; 20: 307-315.
[14] Smyth GK. Limma: linear models for microarray data. Bioinformatics and computational biology solutions using R and Bioconductor. Springer; 2005. pp. 397-420.
[15] Dennis G Jr, Sherman BT, Hosack DA, Yang J, Gao W, Lane HC and Lempicki RA. DAVID: database for annotation, visualization, and integrated discovery. Genome Biol 2003; 4: P3.
[16] Langfelder P and Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics 2008; 9: 559.
[17] Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B and Ideker T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res 2003; 13: 2498-2504.
[18] Franceschini A, Szklarczyk D, Frankild S, Kuhn M, Simonovic M, Roth A, Lin J, Minguez P, Bork P and von Mering C. STRING v9. 1: protein-protein interaction networks, with increased coverage and integration. Nucleic Acids Res 2013; 41: D808-D815.
[19] Quaggin SE, Schwartz L, Cui S, Igarashi P, Deimling J, Post M and Rossant J. The basic-helix-loop-helix protein pod1 is critically important for kidney and lung organogenesis. Development 1999; 126: 5771-5783.
[20] Richards KL, Zhang B, Sun M, Dong W, Churchill J, Bachinski LL, Wilson CD, Baggerly KA, Yin G and Hayes DN. Methylation of the candidate biomarker TCF21 is very frequent across a spectrum of early-stage nonsmall cell lung cancers. Cancer 2011; 117: 606-617.
[21] Acharya A, Baek ST, Banfi S, Eskiocak B and Tallquist MD. Efficient inducible Cre-mediated recombination in Tcf21cell lineages in the heart and kidney. Genesis 2011; 49: 870-877.
[22] Davenport JW, Fernandes ER, Harris LD, Neale GA and Goorha R. The mouse mitotic checkpoint gene bub1b, a novel bub1 family member, is expressed in a cell cycle-dependent manner. Genomics 1999; 55: 113-117.
[23] Tonon G, Wong K-K, Maulik G, Brennan C, Feng B, Zhang Y, Khatry DB, Protopopov A, You MJ and Aguirre AJ. High-resolution genomic profiles of human lung cancer. Proc Natl Acad Sci U S A 2005; 102: 9625-9630.
[24] Kato T, Daigo Y, Aragaki M, Ishikawa K, Sato M and Kaji M. Overexpression of CDC20 predicts poor prognosis in primary non-small cell lung cancer patients. J Surg Oncol 2012; 106: 423-430.
[25] Scintu M, Vitale R, Prencipe M, Gallo AP, Bonghi L, Valori VM, Maiello E, Rinaldi M, Signori E and Rabitti C. Genomic instability and increased expression of BUB1B and MAD2L1 genes in ductal breast carcinoma. Cancer Lett 2007; 254: 298-307.
[26] Abal M, Obrador-Hevia A, Janssen KP, Casadome L, Menendez M, Carpentier S, Barillot E, Wagner M, Ansorge W and Moeslein G. APC inactivation associates with abnormal mitosis completion and concomitant BUB1B/MAD2L1 up-regulation. Gastroenterology 2007; 132: 2448-2458.
[27] Beresford MJ, Wilson GD and Makris A. Measuring proliferation in breast cancer: practicalities and applications. Breast Cancer Res 2006; 8: 216.
[28] Arriola E, Marchio C, Tan DS, Drury SC, Lambros MB, Natrajan R, Rodriguez-Pinilla SM, Mackay A, Tamber N and Fenwick K. Genomic analysis of the HER2/TOP2A amplicon in breast cancer and breast cancer cell lines. Lab Invest 2008; 88: 491-503.
[29] Mueller RE, Parkes RK, Andrulis I and O'Malley FP. Amplification of the TOP2A gene does not predict high levels of topoisomerase II alpha

protein in human breast tumor samples. Genes Chromosomes Cancer 2004; 39: 288-297.

[30] Simon R, Atefy R, Wagner U, Forster T, Fijan A, Bruderer J, Wilber K, Mihatsch MJ, Gasser T and Sauter G. HER-2 and TOP2A coamplification in urinary bladder cancer. Int J Cancer 2003; 107: 764-772.