Integrated analysis reveals candidate genes and transcription factors in lung adenocarcinoma

BAIWANG CHEN*, SHUHONG GAO*, CHANGWEI JI* and GE SONG

Intensive Care Unit, Jining No. 1 People's Hospital, Jining, Shandong 272011, P.R. China

Received February 29, 2016; Accepted February 23, 2017

DOI: 10.3892/mmr.2017.7656

Abstract. Lung adenocarcinoma is the most common type of non-small cell lung cancer in Asia. Therefore, it is important to improve understanding of the underlying transcriptional regulatory mechanisms involved. The present study aimed to identify potential candidate genes and transcription factors (TFs) associated with the disease. Four gene expression profiles were downloaded from the Gene Expression Omnibus database, which included 141 lung adenocarcinoma patients and 191 healthy controls. The differentially expressed genes (DEGs) were screened out and functional annotation was performed. In addition, TFs were identified and a global transcriptional regulatory network was constructed. Integrated analysis gave rise to a total of 1,238 DEGs in lung adenocarcinoma when compared with healthy tissues, including 970 upregulated and 268 downregulated DEGs. The six overexpressed outlier genes of ceruloplasmin, heparan sulfate 6-O-sulfotransferase 2, transmembrane protease serine 4, anillin actin binding protein, cellular retinoic acid binding protein 2 and cystatin SN may serve important roles in the development of lung adenocarcinoma. In addition, the downregulation of carbonic anhydrase 4 and S100 calcium binding protein A12 may render these effective diagnostic biomarkers. The results of the transcriptional regulatory network demonstrated that the hub nodes were sex determining region Y-box 10, Spi-B transcription factor and nuclear receptor subfamily 4 group A member 2. The four TFs, forkhead box D1, E74-like ETS transcription factor 5, homeobox A5 and kruppel-like factor 5, may warrant future investigations into their function in disease development. In conclusion, the present study provided for further studies a list of candidate genes and TFs for the detection and treatment of lung adenocarcinoma.

Correspondence to: Dr Ge Song, Intensive Care Unit, Jining No. 1 People's Hospital, 6 Jiankang Road, Jining, Shandong 272011, P.R. China
E-mail: ge_song_doctor@sina.com

*Contributed equally

Key words: lung adenocarcinoma, candidate genes, transcription factors, integrated analysis

Introduction

Lung adenocarcinoma is a malignant cancer and a primary subtype of non-small cell lung cancer (NSCLC), with the greatest incidence and the worst prognosis worldwide (1). In the majority of cases the development of lung adenocarcinoma is a multifactor and multistage process, which is associated with numerous genes (2).

In addition to the gene expression exhibited by cancer and healthy tissues, identification of the differential interactions between genes in the development of lung adenocarcinoma should be considered, as this may identify critical genes that may not otherwise be detectable (3,4). Transcription factors (TFs) bind to a specific region of the DNA sequence and consequently regulate the transcription of target genes (5,6). Transcriptional regulation is crucial for the development of lung adenocarcinoma (7). Therefore, it is important to construct gene regulatory networks that represent this (8).

Extensive investigation has been performed into the underlying mechanisms of lung adenocarcinoma. A number of studies have assessed gene expression in lung adenocarcinoma (9-12) or identified marker genes (13). Using computational methods, the potential associations between TFs and differentially expressed genes (DEGs) in the regulation of transcription in lung adenocarcinoma have been identified and a regulatory network was constructed (14). A previous study examined the underlying mechanisms of lung adenocarcinoma through the regulatory network using GSE2514 microarray data (7). A previous study on the synergistic regulation of microRNAs (miRNAs) and TFs have identified a variety of significant motifs (15). In addition, a miRNA-TF synergistic regulation network has been constructed (2).

However, the molecular mechanisms underlying lung adenocarcinoma remain to be fully elucidated. Therefore, analysis of the regulatory mechanism based on a large-scale study of genes associated with this disease is important to further the understanding of lung adenocarcinoma. The large body of biological data generated from gene expression profiles is a useful resource for understanding and deducing gene function (16).

The present study performed computational bioinformatics analysis of gene expression for the identification of

potential transcriptional regulation associations between TFs and DEGs in lung adenocarcinoma and adjacent healthy tissue samples. The significantly enriched functions of these genes were investigated to further the understanding of the molecular mechanisms underlying lung adenocarcinoma. In addition, a transcriptional regulatory network was constructed.

Materials and methods

Source of datasets. The transcriptome sequencing data from lung adenocarcinoma patients were downloaded from the Gene Expression Omnibus (GEO) repository (www.ncbi.nlm.nih.gov/geo/) (17). The following key words were used: ['Lung Adenocarcinomas' (MeSH Terms) or 'Lung Adenocarcinomas' (All Fields)] and 'Homo sapiens' (porgn) and 'gse' (Filter). In total, four datasets were obtained with the following accession numbers: GSE62949, GSE27262, GSE43458 and GSE32863. There were 191 cases and 141 controls enrolled in the present study. For each patient, the tumor and paired healthy tissue had been sequenced. The characteristics of the eligible datasets are summarized in Table I.

Differential gene expression analysis. For all datasets, the gene expression level data for cancerous and healthy tissues were preprocessed by background correction and normalization. The limma package (bioconductor.org/packages/release/bioc/html/limma.html) (18) in R was used to analyze differential expression between lung adenocarcinoma and healthy tissues by using a paired Student's t-test. The P-value and false discovery rate (FDR) were obtained. Genes with Benjamini-Hochberg adjusted FDR<0.05 were reported as DEGs in the present study (19).

Gene function annotation and pathway analysis. To assess the alterations in DEGs occurring at the cellular level, and

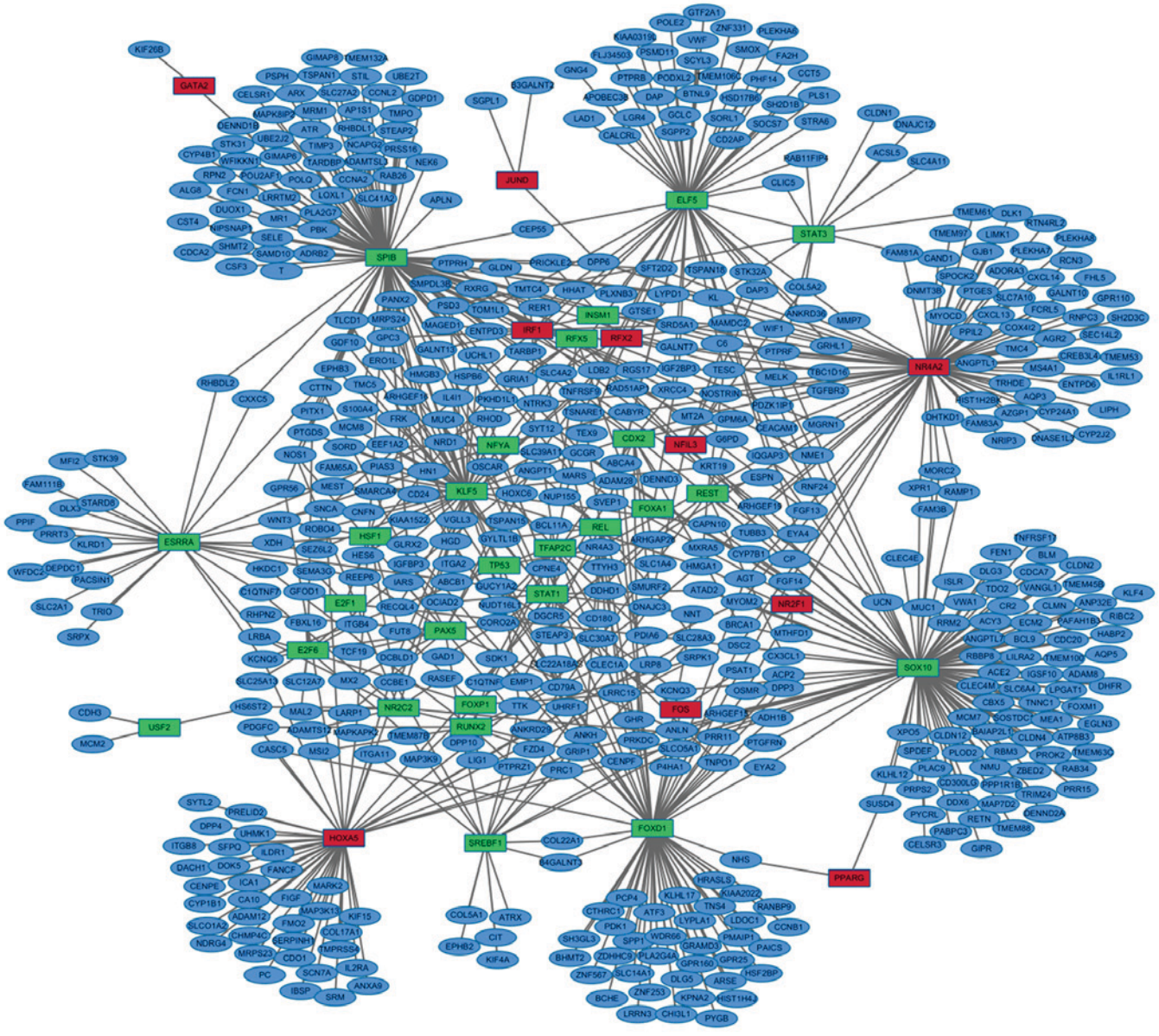

Figure 1. Constructed transcriptional regulatory network of lung adenocarcinoma. Red and green nodes denote upregulated and downregulated transcription factors, respectively. Blue nodes denote target genes.

Table I. Characteristics of the individual GEO studies that produced the eligible datasets used in the present study.

GEO ID	No. of samples (cancer:control)	Platform	Country	Year
GSE62949	28:28	GPL8432 Illumina HumanRef-8 WG-DASL v3.0	USA	2015
GSE27262	25:25	GPL570 [HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array	China (Taiwan)	2013
GSE43458	30:80	GPL6244 [HuGene-1_0-st] Affymetrix Human Gene 1.0 ST Array [transcript (gene) version]	USA	2013
GSE32863	58:58	GPL6480GPL6884 Illumina HumanWG-6 v3.0 expression beadchip	USA	2012

The transcriptome sequencing data from patients with lung adenocarcinoma were downloaded from the GEO database. In total, four datasets were obtained with the following accession numbers: GSE62949, GSE27262, GSE43458 and GSE32863. Collectively, data from 332 individuals were obtained, comprising 141 patients with lung adenocarcinoma and 191 controls. For each patient, the tumor and paired healthy tissue had been sequenced. GEO, Gene Expression Omnibus.

Table II. Primers used in the present study.

Gene	Forward (5'-3')	Reverse (5'-3')	Length (bp)
SLC6A4	CGTGCTCGCCGTGGTCAT	CCCCGTGGCATACTCCTCC	100
SOSTDC1	CCCAGCAGCAACAGCACG	CAGTTCCCGGCAACCCAC	105
TMPRSS4	ACACGGTGCAATGCAGACGA	AGCCATAGCCCCAACTAACGA	169
SOX10	TCAGCGGCTACGACTGGACG	CGTTGTGCAGGTGCGGGTA	156
HOXA5	TTCAACCGTTACCTGACCCGC	TAAACGCTCAGATACTCAGGGACGG	183
β-actin	CTGAAGTACCCCATCGAGCAC	ATAGCACAGCCTGGATAGCAAC	223

the functional clustering of DEGs, the online Gene Ontology Enrichment Analysis and Visualization Tool (cbl-gorilla.cs.technion.ac.il/) (20,21) was used to identify and visualize the Gene Ontology (GO) database categories: Biology process, molecular function and cellular component (22). In addition, GeneCodis3 (genecodis.cnb.csic.es/analysis) (23-25) was used to conduct Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis to investigate the functional roles and associations of the genes with varied expression in the analysis.

Screening of potential TFs. To understand the regulatory mechanisms, the present study further analyzed TFs, which are essential for the regulation of gene activation or repression, in lung adenocarcinoma. The TFs in the human genome and the motifs of genomic binding sites were downloaded from the TRANSFAC® database (http://gene-regulation.com/pub/databases.html) (26), and the DEGs encoding TFs were identified. The TRANSFAC position weight matrix was used for gene promoter scanning to identify DEGs with the TF binding sites in the promoter region. Finally, the transcriptional regulatory networks were established and visualized using Cytoscape 3.0 software (www.cytoscape.org/) (27,28).

RNA preparation and reverse transcription-quantitative polymerase chain reaction (RT-qPCR). A total of 4 patients with lung adenocarcinoma (mean age, 54±1.8) were recruited from Jining No. 1 People's Hospital between October 2015 and December 2015, including two men and two women. Tumors

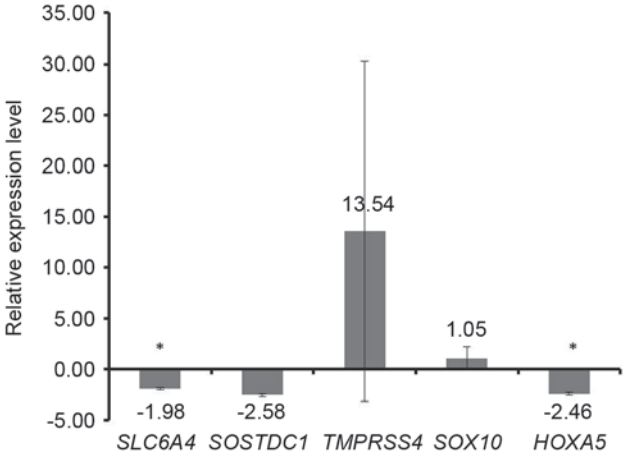

Figure 2. Reverse transcription-quantitative polymerase chain reaction was performed to validate the transcription factors and their targets in the four patients with lung adenocarcinoma compared with the corresponding healthy tissue. Data are presented as the mean ± standard deviation. *P<0.05 vs. corresponding healthy lung tissue (control). SLC6A4, solute carrier family 6 member 4; SOSTDC1, sclerostin domain containing 1; TMPRSS4, transmembrane protease serine 4; SOX10, sex determining region Y-box 10; HOXA5, homeobox A5.

Table III. Top ten upregulated and downregulated differentially expressed genes.

A, Upregulated

ID No.	Symbol	Log FC	FDR
1356	CP	4.10E+00	2.33E-05
9245	GCNT3	3.91E+00	5.76E-05
90161	HS6ST2	6.29E+00	7.22E-04
56649	TMPRSS4	5.27E+00	8.23E-04
54443	ANLN	4.45E+00	1.12E-03
1382	CRABP2	3.93E+00	2.23E-03
9244	CRLF1	3.99E+00	4.52E-03
26585	GREM1	9.27E+00	7.46E-03
1469	CST1	7.34E+00	1.07E-02
7368	UGT8	3.91E+00	4.27E-02

B, Downregulated

ID No.	Symbol	Log FC	FDR
7123	CLEC3B	-4.68128	1.29E-15
762	CA4	-4.78845	1.28E-13
25928	SOSTDC1	-6.45818	4.20E-09
8547	FCN3	-7.00453	9.54E-08
80761	UPK3B	-6.08962	1.09E-06
6532	SLC6A4	-6.64863	1.40E-06
4499	MT1M	-3.70537	2.70E-05
3569	IL6	-3.92147	1.94E-04
6283	S100A12	-4.04627	3.93E-04
9173	IL1RL1	-5.20882	8.34E-03

FC, fold change; FDR, false discovery rate.

and corresponding healthy lung tissue samples were surgically resected, and immediately frozen in liquid nitrogen. Our study was approved by the ethics committee of Jining No. 1 People Hospital. The written informed consents were obtained from each of the patients.

Total RNA was extracted using TRIzol® reagent (Invitrogen; Thermo Fisher Scientific, Inc., Waltham, MA, USA) according to the manufacturer's protocol. RNA was reverse-transcribed using SuperScript® III Reverse Transcriptase (Invitrogen; Thermo Fisher Scientific, Inc.). RT-qPCR was performed using an ABI 7500 Real-Time PCR system (Applied Biosystems; Thermo Fisher Scientific, Inc.) and a Power SYBR® Green PCR Master mix (Invitrogen; Thermo Fisher Scientific, Inc.). qPCR reaction was performed under the following conditions: After denaturing for 10 min at 95˚C, PCR was performed for 45 cycles of 95˚C for 15 sec and 60˚C for 1 min, followed by a 10 min incubation at 72˚C. The results were analyzed using $2^{-\Delta\Delta Cq}$ method (29). Student's t-test was performed to compare the gene expression in cancer and healthy tissue using Microsoft Excel. All reactions were analyzed triplicates. The human β-actin gene was used as endogenous control. The primers were listed in Table II.

Results

Identification of differentially expressed genes. According to the inclusion criteria, four microarray datasets were obtained. Integrated analysis of these generated 1,238 DEGs with FDR<0.05 in lung adenocarcinoma compared with healthy tissues, including 970 upregulated and 268 downregulated DEGs. The top ten up and downregulated DEGs are presented in Table III.

Functional enrichment analysis of DEGs. GO enrichment analysis of DEGs was performed to understand their biological functions. A total of 3 GO categories were investigated: Biological process, cellular component and molecular function. The results revealed that the significantly enriched GO terms for biological process were involved in the regulation of responses to stimulus (GO: 0048583; P=5.00E-07), regulation of defense responses (GO: 0031347; P=2.00E-06) and circulatory system processes (GO: 0003013; P=2.79E-06). In addition, the plasma membrane part (GO: 0044459; P=1.35E-05) was the significantly enriched GO term for cellular component. Notably, the significantly enriched GO term for molecular function was receptor activity (GO: 0004872; P=6.43E-07; Table IV).

KEGG pathway enrichment analysis indicated that extracellular matrix (ECM)-receptor interaction (FDR=1.05E-07) and cell cycle (FDR=1.18E-07) were significantly enriched. Furthermore, focal adhesion (FDR=9.43E-06), cell adhesion molecules (FDR=7.72E-05) and pathways in cancer (FDR=1.71E-04) were enriched (Table V).

Construction of TF-target gene regulatory network for lung adenocarcinoma. To construct the TF-target gene regulatory network for lung adenocarcinoma, the TRANSFAC database was utilized to investigate TFs and their latent target genes. Differentially expressed TFs and latent target genes in lung adenocarcinoma were selected. A total of 40 differentially expressed TFs (27 upregulated and 13 downregulated) and 544 latent differentially expressed target genes were identified. The transcriptional regulatory network was subsequently constructed based on these findings. In the network, there were 36 TFs and 752 TF-target interactions (Fig. 1). The top ten TFs regulating the greatest number of downstream target genes were sex determining region Y-box 10 (SOX10), Spi-B transcription factor (SPIB), nuclear receptor subfamily 4 group A member 2 (NR4A2), forkhead box D1 (FOXD1), E74 like ETS transcription factor 5 (ELF5), homeobox A5 (HOXA5), kruppel like factor 5 (KLF5), estrogen related receptor α (ESRRA), sterol regulatory element binding transcription factor 1 (SREBF1) and REL proto-oncogene, NF-kB subunit (REL; Table VI).

Validation of differentially expressed TFs and targets. Tumor and corresponding healthy lung tissue samples were used to validate the findings of the integrated analysis. Two TFs of SOX10 and HOXA5 were selected, where SOX10 had the highest number of downstream DEGs and HOXA5

Table IV. Enriched Gene Ontology database terms of differentially expressed genes.

A, Biological process

GO ID	GO term	No. of genes	P-value
GO:0048583	Regulation of response to stimulus	66	5.00E-07
GO:0071310	Cellular response to organic substance	42	8.33E-07
GO:0070887	Cellular response to chemical stimulus	49	1.36E-06
GO:0032501	Multicellular organismal process	89	1.60E-06
GO:0031347	Regulation of defense response	31	2.00E-06
GO:0003013	Circulatory system process	14	2.79E-06
GO:0003008	System process	43	2.99E-06
GO:0050729	Positive regulation of inflammatory response	6	3.22E-06
GO:0007186	G-protein coupled receptor signaling pathway	32	5.06E-06
GO:1903034	Regulation of response to wounding	30	5.32E-06
GO:0044707	Single-multicellular organism process	87	5.80E-06
GO:0010033	Response to organic substance	65	6.71E-06
GO:0033993	Response to lipid	42	9.57E-06
GO:0002682	Regulation of immune system process	40	9.58E-06
GO:0042221	Response to chemical	75	1.04E-05

B, Molecular function

GO ID	GO term	No. of genes	P-value
GO:0004872	Receptor activity	55	6.43E-07
GO:0038023	Signaling receptor activity	46	1.62E-06
GO:0060089	Molecular transducer activity	58	5.70E-06
GO:0004871	Signal transducer activity	51	2.22E-05
GO:0004888	Transmembrane signaling receptor activity	40	6.35E-05
GO:0005102	Receptor binding	42	2.57E-04
GO:0004908	Interleukin-1 receptor activity	2	4.14E-04
GO:0050998	Nitric-oxide synthase binding	2	5.68E-04
GO:0038187	Pattern recognition receptor activity	3	7.47E-04
GO:0008329	Signaling pattern recognition receptor activity	3	7.47E-04
GO:0050431	Transforming growth factor beta binding	5	9.29E-04

C, Cellular component

GO ID	GO term	No. of genes	P-value
GO:0044459	Plasma membrane part	79	1.35E-05
GO:0005886	Plasma membrane	89	1.12E-04
GO:0031226	Intrinsic component of plasma membrane	52	1.38E-04
GO:0005887	Integral component of plasma membrane	48	3.53E-04
GO:0031526	Brush border membrane	3	5.25E-04
GO:0005576	Extracellular region	47	5.43E-04

GO, Gene Ontology database.

was the primary significantly upregulated TF. In addition, solute carrier family 6 member 4 (SLC6A4) and sclerostin domain containing 1 (SOSTDC1) were two targets of SOX10 and they were listed in the top 50 significant DEGs. The transmembrane protease serine 4 (TMPRSS4) was a target of HOXA5 and it was listed in the top 300 significant DEGs. Therefore, SLC6A4, SOSTDC1 and TMPRSS4 were selected for validation. The RT-qPCR results demonstrated that the

Table V. Top 15 enriched KEGG pathways of differentially expressed genes.

KEGG ID	KEGG term	Count	FDR
hsa04512	ECM-receptor interaction	18	1.05E-07
hsa04110	Cell cycle	22	1.18E-07
hsa04115	p53 signaling pathway	14	5.46E-06
hsa04510	Focal adhesion	24	9.43E-06
hsa04514	Cell adhesion molecules	9	7.72E-05
hsa04670	Leukocyte transendothelial migration	9	7.72E-05
hsa03030	DNA replication	9	8.80E-05
hsa04974	Protein digestion and absorption	13	1.04E-04
hsa05200	Pathways in cancer	29	1.71E-04
hsa00512	Mucin type O-Glycan biosynthesis	8	1.71E-04
hsa04114	Oocyte meiosis	9	1.93E-04
hsa00250	Alanine, aspartate and glutamate metabolism	8	2.44E-04
hsa04530	Tight junction	7	2.63E-04
hsa04614	Renin-angiotensin system	6	2.99E-04
hsa04060	Cytokine-cytokine receptor interaction	24	3.43E-04

KEGG, Kyoto Encyclopedia of Genes and Genomes; FDR, false discovery rate; ECM, extracellular matrix.

expression pattern of the selected genes was similar to that identified by the integrated analysis. SOX10 and TMPRSS4 were upregulated, whereas SLC6A4, SOSTDC1 and HOXA5 were downregulated in lung adenocarcinoma compared with the corresponding healthy lung tissue samples (Fig. 2). The significance of difference was slightly different, which was primarily due to the difference of sample number. There were 4 pairs samples used in the RT-qPCR experiment, whereas there were 191 cases and 141 controls in the integrated analysis.

Discussion

Lung adenocarcinoma is the most common histological subtype of lung cancer. The present study investigated the molecular mechanisms underlying lung adenocarcinoma through the regulatory network using microarray datasets obtained from the GEO database. Integrated analysis of four microarray datasets identified a total of 1,238 DEGs (970 upregulated and 268 downregulated) in lung adenocarcinoma compared with healthy tissues. Functional annotation demonstrated that DEGs were closely associated with common pathways for cancers, including the cell cycle, p53 signaling pathway and pathways in cancer. In addition, ECM-receptor interactions, focal adhesion and cell adhesion molecules were significantly enriched, which may be closely associated with tumorigenesis in lung adenocarcinoma.

Of the top ten upregulated and downregulated DEGs, the majority were associated with the pathological process of lung adenocarcinoma. A previous study demonstrated that ceruloplasmin (CP) was overexpressed at a high frequency in lung adenocarcinoma compared with corresponding healthy lung tissues (30). Heparan sulfate 6-O-sulfotransferase 2 (HS6ST2) is significantly overexpressed in lung tumor tissues (31). TMPRSS4 expression has been associated with postoperative recurrence in patients with lung cancer (32). In addition, anillin actin binding protein (ANLN) has been reported to be essential for the formation or organization of actin cables in the cleavage furrow and serves an important role in cytokinesis (33). Suzuki *et al* (34) demonstrated that ANLN was overexpressed in the majority of primary NSCLCs, and the endogenous expression of ANLN in the nucleus was significantly associated with poor prognosis in patients with NSCLC. Cellular retinoic acid binding protein 2 (CRABP2) expression is markedly increased in lung adenocarcinoma (35) and the expression of cystatin SN (CST1) is closely associated with tumor metastasis properties in A549L6 cells (36). In the present study, CP, HS6ST2, TMPRSS4, ANLN, CRABP2 and CST1 were significantly upregulated in lung adenocarcinoma compared with corresponding healthy lung tissues, indicating that the overexpressed outlier genes may serve important roles in the development of lung adenocarcinoma.

In addition, three of the top ten downregulated genes have been reported previously. Carbonic anhydrase 4 is downregulated in lung adenocarcinoma (37). S100 calcium binding protein A12 (S100A12) is a proinflammatory marker that has the potential to be a diagnostic biomarker of NSCLC (38). A recent study reported that interleukin (IL)-6 is upregulated in lung adenocarcinoma and suggested that IL-6 may be a therapeutic target for the treatment of V-Ki-ras2 Kirsten rat sarcoma viral oncogene homolog-driven lung adenocarcinoma (39). However, the present study revealed that IL-6 was downregulated in lung adenocarcinoma. Therefore, further studies are required.

Using the TRANSFAC database, 40 differentially expressed TFs were identified in lung adenocarcinoma and a transcriptional regulatory network was constructed. Based on the constructed transcriptional regulatory network, a set of crucial TFs, which had the highest number of downstream DEGs, were identified as being of interest, including SOX10,

Table VI. Top ten transcription factors interacting with the greatest number of differentially expressed genes.

TF	P-value	Up/down	Count	Genes
SOX10	9.43E-02	Up	109	ANP32E, RIBC2, GRIP1, ADH1B, RETN, FEN1, HOXA5, PAFAH1B3, MELK, DDX6, BAIAP2L1, SYT12, TNNC1, DHFR, CLMN, PLAC9, FOXM1, CDC20, FAM3B, SLC6A4, CELSR3, KLHL12, ANGPTL7, TMEM100, TDO2, EYA2, IGSF10, SOSTDC1, ANLN, EGLN3, CLEC4E, RAMP1, ADAM8, CAPN10, KRT19, RBM3, C6, PABPC3, ARHGEF15, MEA1, KLF4, DENND2A, SPDEF, AQP5, TRIM24, RAD51AP1, GCGR, CBX5, VANGL1, SLCO5A1, IQGAP3, MARS, PTPRF, LILRA2, MYOM2, CD300LG, ECM2, CLDN12, KCNQ3, LPGAT1, CLEC4M, CX3CL1, CR2, HABP2, ATP8B3, MCM7, ATAD2, PTGFRN, CYP7B1, CDCA7, MAP7D2, DLG3, MUC1, RAB34, RRM2, CLDN4, GIPR, DPP3, ISLR, CENPF, ACP2, MORC2, PDIA6, UCN, XPO5, SUSD4, ACY3, PRPS2, BLM, TMEM45B, RBBP8, TNFRSF17, TMEM63C, VWA1, XPR1, PPP1R1B, PROK2, BRCA1, TMEM88, HMGA1, ACE2, PLOD2, TNFRSF9, PRR15, NMU, BCL9, PYCRL, ZBED2, CLDN2
SPIB	6.34E-02	Up	108	PANX2, PTPRF, CEP55, TLCD1, PBK, NIPSNAP1, ADAMTSL3, ARHGEF19, TSPAN15, ENTPD3, ATR, CSF3, GALNT7, HGD, UBE2T, NOS1, MR1, SLC27A2, CDCA2, T, GIMAP8, MAGED1, TOM1L1, PTPRH, PLXNB3, LYPD1, POLQ, RGS17, DPP6, HMGB3, SHMT2, CST4, CYP4B1, CTTN, FRK, ALG8, KL, PITX1, TARDBP, ERO1L, NEK6, CCNL2, WFIKKN1, RXRG, PSD3, RPN2, RAB26, RER1, PLA2G7, C6, GDF10, SMPDL3B, MRPS24, CELSR1, RHBDL2, DUOX1, PDZK1IP1, TIMP3, MRM1, ADRB2, TSPAN1, LOXL1, APLN, IGFBP3, PRSS16, GLRX2, NME1, SYT12, GALNT13, TMC5, RAD51AP1, POU2AF1, SAMD10, SFT2D2, UBE2J2, RHBDL1, SELE, MAPK8IP2, TMEM132A, GATA2, HN1, STEAP2, MAPKAPK2, FCN1, PSPH, MGRN1, TMPO, GPC3, GDPD1, GPR56, MARS, LRRTM2, GUCY1A2, GIMAP6, SLC41A2, CEACAM1, STK31, TNFRSF9, CCNA2, NCAPG2, HOXC6, AP1S1, DENND1B, GCGR, STIL, ARX, CD24, CXXC5
NR4A2	1.45E-02	Down	81	CABYR, AQP3, RTN4RL2, MORC2, DPP6, SEC14L2, CAND1, MMP7, TRHDE, PLEKHA8, ANGPTL1, DHTKD1, PLEKHA7, MAMDC2, NRIP3, FAM83A, ARHGAP26, LIPH, TMC4, RGS17, ENTPD6, NME1, GALNT10, PTPRF, MYOM2, LYPD1, DNAJC3, COX4I2, DNMT3B, GALNT7, AGR2, GJB1, TMEM97, MGRN1, AGT, SRD5A1, LIMK1, TMEM61, CEACAM1, MT2A, CYP2J2, SFT2D2, GRHL1, DSC2, PTGES, FHL5, CYP24A1, TMEM53, FAM81A, CREB3L4, OSMR, SLC7A10, CXCL13, DNASE1L3, GPR110, PDZK1IP1, AZGP1, RAMP1, FAM3B, SPOCK2, RCN3, MS4A1, NR4A3, C6, CXCL14, ADORA3, HIST1H2BK, MYOCD, ANKRD36, WIF1, RNPC3, NOSTRIN, DLK1, FCRL5, IL1RL1, PPIL2, FGF14, XPR1, NTRK3, SH2D3C, ARHGEF19
FOXD1	8.42E-02	Up	68	PAICS, TNPO1, PRR11, HRASLS, ZNF253, DSC2, COL22A1, EYA2, PRKDC, ZDHHC9, DDHD1, HSF2BP, DPP3, SPP1, ABCA4, BHMT2, ADH1B, AGT, PLA2G4A, DLG5, RANBP9, UHRF1, CLEC1A, TNS4, ADAM28, LDOC1, ANLN, NNT, ZNF567, CCNB1, GUCY1A2, NHS, ARSE, CTHRC1, CD24, PTGFRN, CHI3L1, CASC5, PYGB, ATF3, PMAIP1, HOXC6, HIST1H4J, LYPLA1, GPR25, KIAA2022, PRC1, HN1, GLRX2, IGFBP3, TSPAN15, PDK1, WDR66, ACP2, KPNA2, PCP4, B4GALNT3, SLC14A1, GPR160, HGD, FGF14, P4HA1, SH3GL3, OSMR, GRAMD3, LRRN3, KLHL17, BCHE

Table VI. Continued.

TF	P-value	Up/down	Count	Genes
ELF5	5.48E-02	Up	59	PTPRB, RNF24, POLE2, PLXNB3, ANKRD36, TMEM106C, CLIC5, FA2H, APOBEC3B, SGPP2, GNG4, ZNF331, TSPAN18, DAP3, SORL1, FAM81A, TOM1L1, STRA6, LDB2, DAP, GTF2A1, VGLL3, LAD1, BTNL9, MMP7, PHF14, TESC, ABCA4, XRCC4, TARBP1, SMOX, SOCS7, GRHL1, STK32A, PLS1, PODXL2, VWF, TEX9, IGF2BP3, CD2AP, KIAA0319L, FLJ34503, PSMD11, CCT5, GCLC, CEP55, RAB11FIP4, LGR4, SNCA, IQGAP3, HSD17B6, SCYL3, CALCRL, EYA4, TSNARE1, PLEKHA6, RER1, SH2D1B, ADAM28
HOXA5	4.13E-04	Down	50	NR4A3, CENPF, CENPE, SERPINH1, CYP1B1, IBSP, ICA1, IL2RA, ITGB8, SNCA, SFPQ, ITGA11, CDO1, NDRG4, FANCF, SYTL2, VGLL3, PRELID2, MRPS23, FMO2, DNAJC3, ILDR1, DACH1, SLC28A3, CHMP4C, ADAM12, DCBLD1, TMPRSS4, ANKRD29, MAP3K13, KIF15, SRPK1, DOK5, COL17A1, MARK2, PC, SRM, CASC5, DPP4, ANXA9, PDGFC, SCN7A, CA10, SLCO1A2, FIGF, KCNQ5, UHMK1, GPR56, GRIP1, MAPKAPK2
KLF5	9.34E-02	Up	45	SORD,GPR56,PSD3,HMGA1,SMARCA4,CNFN,TLCD1,PKHD1L1, GYLTL1B, SEZ6L2, TNFRSF9, HSPB6, PTGDS, RECQL4, KRT19, MUC4, SLC4A2, ROBO4, ABCB1, DCBLD1, EEF1A2, PANX2, UHRF1, LDB2, PIAS3, MT2A, CAPN10, FAM65A, EPHB3, S100A4, SLC39A11, GLRX2, MEST, GDF10, UCHL1, GRIA1, HMGB3, HES6, RHOD, ARHGEF16, IL4I1, NR2C2, MRPS24, MCM8, KIAA1522
ESRRA	7.63E-02	Up	27	WFDC2, MFI2, PACSIN1, C1QTNF7, XDH, KCNQ5, RHBDL2, EPHB3, DEPDC1, STK39, CXXC5, SLC2A1, PRRT3, HKDC1, PPIF, STARD8, KLRD1, DLX3, SRPX, SNCA, TRIO, NNT, SEZ6L2, FAM111B, LRBA, NTRK3, REEP6
SREBF1	5.74E-02	Up	18	COL22A1,SLC28A3,DPP10,CIT,ATRX,EPHB2,CCBE1,TMEM87B, DNAJC3, ITGA11, KIF4A, B4GALNT3, TTK, PTPRZ1, ARHGEF15, COL5A1, ANKH, SLCO5A1
REL	2.16E-02	Up	14	CLEC1A, NUP155, LRP8, ANGPT1, ANKH, GHR, KCNQ5, EMP1, FOXA1, TEX9, SLC28A3, SLC30A7, GYLTL1B, LRRC15

Up, upregulated; down, downregulated; SOX10, sex determining region Y-box 10; SPIB, Spi-B transcription factor; NR4A2, nuclear receptor subfamily 4 group A member 2; FOXD1, forkhead box D1; ELF5, E74 like ETS transcription factor 5; HOXA5, homeobox A5; KLF5, kruppel like factor 5; ESRRA, estrogen related receptor α; SREBF1, sterol regulatory element binding transcription factor 1; REL, REL proto-oncogene, NF-κB subunit; TF, transcription factor.

SPIB, NR4A2, FOXD1, ELF5, HOXA5, KLF5, ESRRA, SREBF1 and REL. A number of TFs may serve important roles in the development of lung adenocarcinoma.

The human Forkhead-box gene family consists of at least 43 members (40), including FOXD1, the loss of which may suppress cell proliferation and significantly increase the life expectancy of patients with NSCLC (41). A study using a transgenic mouse model of papillary lung adenocarcinomas revealed that ELF5 may cooperate with c-Myc to suppress and upregulate genes in cancer samples, which may serve an essential role in neoplastic transformation (42). Microarrays of invasion/metastasis lung adenocarcinoma cell lines revealed that HOXA5 may contribute to the suppression of metastasis in lung cancer via the regulation of cytoskeleton remodeling. In addition, KLF5 may promote the apoptosis of lung adenocarcinoma cells, potentially via the inhibition of cell proliferation and repair/activation of apoptosis pathway proteins (43). Therefore, the results of the present study may be useful for future investigations into the role of transcription factors in the development of this complex disease.

In conclusion, the present study generated a list of candidate genes and TFs for the future detection and treatment of lung adenocarcinoma. The results highlighted the potential mechanisms underlying human lung adenocarcinoma through the transcriptional regulatory network.

References

1. Minna JD, Roth JA and Gazdar AF: Focus on lung cancer. Cancer cell 1: 49-52, 2002.

2. Zhao N, Liu Y, Chang Z, Li K, Zhang R, Zhou Y, Qiu F, Han X and Xu Y: Identification of biomarker and co-regulatory motifs in lung adenocarcinoma based on differential interactions. PLoS One 10: e0139165, 2015.
3. Bandyopadhyay S, Mehta M, Kuo D, Sung MK, Chuang R, Jaehnig EJ, Bodenmiller B, Licon K, Copeland W, Shales M, *et al*: Rewiring of genetic networks in response to DNA damage. Science 330: 1385-1389, 2010.
4. Liu X, Liu ZP, Zhao XM and Chen L: Identifying disease genes and module biomarkers by differential interactions. J Am Med Inform Assoc 19: 241-248, 2012.
5. Latchman DS: Transcription factors: An overview. Int J Biochem Cell Biol 29: 1305-1312, 1997.
6. Mitchell PJ and Tjian R: Transcriptional regulation in mammalian cells by sequence-specific DNA binding proteins. Science 245: 371-378, 1989.
7. Meng X, Lu P, Bai H, Xiao P and Fan Q: Transcriptional regulatory networks in human lung adenocarcinoma. Mol Med Rep 6: 961-966, 2012.
8. Chen CY, Chen ST, Fuh CS, Juan HF and Huang HC: Coregulation of transcription factors and microRNAs in human transcriptional regulatory network. BMC Bioinformatics 12 (Suppl 1): S41, 2011.
9. Garber ME, Troyanskaya OG, Schluens K, Petersen S, Thaesler Z, Pacyna-Gengelbach M, van de Rijn M, Rosen GD, Perou CM, Whyte RI, *et al*: Diversity of gene expression in adenocarcinoma of the lung. Proc Natl Acad Sci USA 98: 13784-13789, 2001.
10. Director's Challenge Consortium for the Molecular Classification of Lung Adenocarcinoma, Shedden K, Taylor JM, Enkemann SA, Tsao MS, Yeatman TJ, Gerald WL, Eschrich S, Jurisica I, Giordano TJ, *et al*: Gene expression-based survival prediction in lung adenocarcinoma: A multi-site, blinded validation study. Nat Med 14: 822-827, 2008.
11. Stearman RS, Dwyer-Nield L, Zerbe L, Blaine SA, Chan Z, Bunn PA Jr, Johnson GL, Hirsch FR, Merrick DT, Franklin WA, *et al*: Analysis of orthologous gene expression between human pulmonary adenocarcinoma and a carcinogen-induced murine model. Am J Pathol 167: 1763-1775, 2005.
12. Zhang W, Gong W, Ai H, Tang J and Shen C: Gene expression analysis of lung adenocarcinoma and matched adjacent non-tumor lung tissue. Tumori 100: 338-345, 2014.
13. Jiang H, Deng Y, Chen HS, Tao L, Sha Q, Chen J, Tsai CJ and Zhang S: Joint analysis of two microarray gene-expression data sets to select lung adenocarcinoma marker genes. BMC Bioinformatics 5: 81, 2004.
14. Fu S, Pan X and Fang W: Differential co-expression analysis of a microarray gene expression profiles of pulmonary adenocarcinoma. Mol Med Rep 10: 713-718, 2014.
15. Lin CC, Chen YJ, Chen CY, Oyang YJ, Juan HF and Huang HC: Crosstalk between transcription factors and microRNAs in human protein interaction network. BMC Syst Biol 6: 18, 2012.
16. Li BQ, You J, Chen L, Zhang J, Zhang N, Li HP, Huang T, Kong XY and Cai YD: Identification of lung-cancer-related genes with the shortest path approach in a protein-protein interaction network. Biomed Res Int 2013: 267375, 2013.
17. Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, Marshall KA, Phillippy KH, Sherman PM, Holko M, *et al*: NCBI GEO: Archive for functional genomics data sets-update. Nucleic Acids Res 41 (Database Issue): D991-D995, 2013.
18. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W and Smyth GK: limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res 43: e47, 2015.
19. Hardcastle TJ: Generalized empirical Bayesian methods for discovery of differential data in high-throughput biology. Bioinformatics 32: 195-202, 2016.
20. Eden E, Navon R, Steinfeld I, Lipson D and Yakhini Z: GOrilla: A tool for discovery and visualization of enriched GO terms in ranked gene lists. BMC Bioinformatics 10: 48, 2009.
21. Eden E, Lipson D, Yogev S and Yakhini Z: Discovering motifs in ranked lists of DNA sequences. PLoS Comput Biol 3: e39, 2007.
22. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, *et al*: Gene ontology: Tool for the unification of biology. The gene ontology consortium. Nat Genet 25: 25-29, 2000.
23. Nogales-Cadenas R, Carmona-Saez P, Vazquez M, Vicente C, Yang X, Tirado F, Carazo JM and Pascual-Montano A: GeneCodis: Interpreting gene lists through enrichment analysis and integration of diverse biological information. Nucleic Acids Res 37 (Web Server Issue): W317-W322, 2009.
24. Tabas-Madrid D, Nogales-Cadenas R and Pascual-Montano A: GeneCodis3: A non-redundant and modular enrichment analysis tool for functional genomics. Nucleic Acids Res 40 (Web Server Issue): W478-W483, 2012.
25. Carmona-Saez P, Chagoyen M, Tirado F, Carazo JM and Pascual-Montano A: GENECODIS: A web-based tool for finding significant concurrent annotations in gene lists. Genome Biol 8: R3, 2007.
26. Matys V, Fricke E, Geffers R, Gössling E, Haubrock M, Hehl R, Hornischer K, Karas D, Kel AE, Kel-Margoulis OV, *et al*: TRANSFAC: Transcriptional regulation, from patterns to profiles. Nucleic Acids Res 31: 374-378, 2003.
27. Smoot ME, Ono K, Ruscheinski J, Wang PL and Ideker T: Cytoscape 2.8: New features for data integration and network visualization. Bioinformatics 27: 431-432, 2011.
28. 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 13: 2498-2504, 2003.
29. Livak KJ and Schmittgen TD: Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) method. Methods 25: 402-408, 2001.
30. Wang KK, Liu N, Radulovich N, Wigle DA, Johnston MR, Shepherd FA, Minden MD and Tsao MS: Novel candidate tumor marker genes for lung adenocarcinoma. Oncogene 21: 7598-7604, 2002.
31. Backen AC, Cole CL, Lau SC, Clamp AR, McVey R, Gallagher JT and Jayson GC: Heparan sulphate synthetic and editing enzymes in ovarian cancer. Br J Cancer 96: 1544-1548, 2007.
32. Chikaishi Y, Uramoto H, Koyanagi Y, Yamada S, Yano S and Tanaka F: TMPRSS4 expression as a marker of recurrence in patients with lung cancer. Anticancer Res 36: 121-127, 2016.
33. Oegema K, Savoian MS, Mitchison TJ and Field CM: Functional analysis of a human homologue of the Drosophila actin binding protein anillin suggests a role in cytokinesis. J Cell Biol 150: 539-552, 2000.
34. Suzuki C, Daigo Y, Ishikawa N, Kato T, Hayama S, Ito T, Tsuchiya E and Nakamura Y: ANLN plays a critical role in human lung carcinogenesis through the activation of RHOA and by involvement in the phosphoinositide 3-kinase/AKT pathway. Cancer Res 65: 11314-11325, 2005.
35. Han SS, Kim WJ, Hong Y, Hong SH, Lee SJ, Ryu DR, Lee W, Cho YH, Lee S, Ryu YJ, *et al*: RNA sequencing identifies novel markers of non-small cell lung cancer. Lung Cancer 84: 229-235, 2014.
36. Cai X, Luo J, Yang X, Deng H, Zhang J, Li S, Wei H, Yang C, Xu L, Jin R, *et al*: In vivo selection for spine-derived highly metastatic lung cancer cells is associated with increased migration, inflammation and decreased adhesion. Oncotarget 6: 22905-22917, 2015.
37. Chen L, Zhuo D, Chen J and Yuan H: Screening feature genes of lung carcinoma with DNA microarray analysis. Int J Clin Exp Med 8: 12161-12171, 2015.
38. Lim MY and Thomas PS: Biomarkers in exhaled breath condensate and serum of chronic obstructive pulmonary disease and non-small-cell lung cancer. Int J Chronic Dis 2013: 578613, 2013.
39. Brooks GD, McLeod L, Alhayyani S, Miller A, Russell PA, Ferlin W, Rose-John S, Ruwanpura S and Jenkins BJ: IL6 Trans-signaling Promotes KRAS-Driven lung carcinogenesis. Cancer Res 76: 866-876, 2016.
40. Katoh M and Katoh M: Human FOX gene family (Review). Int J Oncol 25: 1495-1500, 2004.
41. Nakayama S, Soejima K, Yasuda H, Yoda S, Satomi R, Ikemura S, Terai H, Sato T, Yamaguchi N, Hamamoto J, *et al*: FOXD1 expression is associated with poor prognosis in non-small cell lung cancer. Anticancer Res 35: 261-268, 2015.
42. Ciribilli Y, Singh P, Spanel R, Inga A and Borlak J: Decoding c-Myc networks of cell cycle and apoptosis regulated genes in a transgenic mouse model of papillary lung adenocarcinomas. Oncotarget 6: 31569-31592, 2015.
43. Wang CC, Su KY, Chen HY, Chang SY, Shen CF, Hsieh CH, Hong QS, Chiang CC, Chang GC, Yu SL and Chen JJ: HOXA5 inhibits metastasis via regulating cytoskeletal remodelling and associates with prolonged survival in non-small-cell lung carcinoma. PLoS One 10: e0124191, 2015.