Differential co-expression analysis of a microarray gene expression profiles of pulmonary adenocarcinoma

SHIJIE FU, XUFENG PAN and WENTAO FANG

Department of Thoracic Surgery, Shanghai Chest Hospital, Shanghai Jiaotong University, Shanghai 200030, P.R. China

Received September 2, 2013; Accepted March 17, 2014

DOI: 10.3892/mmr.2014.2300

Abstract. Lung cancer severely reduces the quality of life worldwide and causes high socioeconomic burdens. However, key genes leading to the generation of pulmonary adenocarcinoma remain elusive despite intensive research efforts. The present study aimed to identify the potential associations between transcription factors (TFs) and differentially co-expressed genes (DCGs) in the regulation of transcription in pulmonary adenocarcinoma. Gene expression profiles of pulmonary adenocarcinoma were downloaded from the Gene Expression Omnibus, and gene expression was analyzed using a computational method. A total of 37,094 differentially co-expressed links (DCLs) and 251 DCGs were identified, which were significantly enriched in 10 pathways. The construction of the regulatory network and the analysis of the regulatory impact factors revealed eight crucial TFs in the regulatory network. These TFs regulated the expression of DCGs by promoting or inhibiting their expression. In addition, certain TFs and target genes associated with DCGs did not appear in the DCLs, which indicated that those TFs could be synergistic with other factors. This is likely to provide novel insights for research into pulmonary adenocarcinoma. In conclusion, the present study may enhance the understanding of disease mechanisms and lead to an improved diagnosis of lung cancer. However, further studies are required to confirm these observations.

Introduction

Globally, lung cancer is the leading cause of cancer mortality in males and the second leading cause of cancer mortality in females with ~1.6 million new lung cancer cases and 1.4 million mortalities expected to occur in one year (1). In China, the most frequently diagnosed cancer is lung cancer (2), and Chinese females have higher lung cancer rates than females in several European countries (3).

Almost 40% of lung cancer cases are adenocarcinoma, which usually originates in the peripheral lung tissue (4). Although several studies have assessed gene expression (5-7) or provided novel diagnostic aids (8) for pulmonary adenocarcinoma, key genes leading to the deterioration of pulmonary adenocarcinoma remain to be fully evaluated. Extensive efforts have been made to uncover the basic mechanisms underlying the initiation and progression of pulmonary adenocarcinoma, as well as to target these processes for diagnostics at molecular and genetic levels. Key genes and transcription factors (TFs) have an important role in the study of the disease. For example, the expression of the receptor of activated C kinase, which is an important 36-kDa cytosolic protein (9), was reported to be a useful biomarker for pulmonary adenocarcinoma (10). Furthermore, thyroid transcription factor-1 gene amplification has been discovered in certain types of lung adenocarcinoma, and this finding has been useful to inhibit transforming growth factor-β-mediated epithelial-to-mesenchymal transition in pulmonary adenocarcinoma cells (11). Compared with traditional research methods, DNA microarrays are one of the most popular technologies for studying the expression of genes at a large scale and ultimately associating them with diseases (12). The molecular mechanisms of pulmonary adenocarcinoma have yet to be fully understood, and a large-scale study of genes associated with this disease is necessary.

The aim of the present study was to explore the biochemical pathways leading to the deterioration of patients with pulmonary adenocarcinoma at the gene transcription level using a computational bioinformatics analysis of gene expression. Furthermore, the study aimed to identify the potential association between TFs and differentially co-expressed genes (DCGs) in the regulation of transcription. The present study may provide the groundwork to enable the exploration of the most variable genes leading to pulmonary adenocarcinoma.

Materials and methods

All patients provided informed consent prior to their inclusion in the present study, and all human studies were approved by the Ethics Committee of Shanghai Jiaotong University (Shanghai, China) and performed in accordance with the ethical standards.

Correspondence to: Dr Wentao Fang, Department of Thoracic Surgery, Shanghai Chest Hospital, Shanghai Jiaotong University, No. 241 Huaihai East Road, Xuhui, Shanghai 200030, P.R. China
E-mail: chest2013@126.com

Key words: pulmonary adenocarcinoma, differentially co-expressed genes, pathway

Affymetrix microarray data and differential expression analysis. The gene expression profile of GSE 2514 (7) was downloaded from a public functional genomics data repository,
the Gene Expression Omnibus (GEO), which was based on the Affymetrix GPL8300 platform (Affymetrix Human Genome U95, Version 2 Array; Affymetrix, Inc., Santa Clara, CA, USA). These expression data were deposited by Stearman et al (7). A total of 39 samples, including 20 human lung cancer tissue samples and 19 human normal lung tissue samples, were analyzed with one replicate each.

The R package was used to analyze the gene expression profile (http://r-project.org/). The CEL source files were processed into expression estimates, and background correction and quartile data normalization were performed using the Robust Multi-array Average algorithm (13). The probability of genes being differentially expressed between pulmonary adenocarcinoma samples and normal samples was computed using the limma package (14). The t-test method was used to identify DEGs (15,16). P-values <0.05 and |logFC|>0.5 were considered to be statistically significant. The DCsum, DCp and DCe functions in the Differential Co-expression Analysis and Differential Regulation Analysis of Gene Expression Microarray Data (DCGL) (17,18) (part of the R package) were used to evaluate DCGs and differentially co-expressed links (DCLs).

A Q-value <0.05 was defined as the cut-off criterion.

Pathway enrichment analysis of DCGs. The Kyoto Encyclopedia of Genes and Genomes (KEGG) PATHWAY database is a comprehensive database containing various biochemical pathways (19). This database records networks of molecular interactions in cells, as well as variants of these networks specific to particular organisms. To explore the dysfunctional pathways in pulmonary adenocarcinoma samples, each group of genes was assessed using the Database for Annotation, Visualization and Integrated Discovery (DAVID) (20) for pathway enrichment analysis. The DAVID is a program that detects an enrichment of genes with specific gene ontology, KEGG and SwissProt terms.

Construction of a transcriptional regulatory network. TRANSFAC® (21) is a database of TFs, their genomic binding sites and DNA-binding profiles. TRANSFAC® comprises numerous data sheets, including SITE, GENE, FACTOR, CLASS, MATRIX, CELLS, METHOD and REFERENCE. The association between downloaded human TFs and target genes in the database was analyzed. A total of 298 TFs and 6,495 TF-target pairs were selected. The DCLs were mapped to the TF-target pairs and the results were associated with the known target genes to obtain transcriptional regulation interrelations. Cytoscape (22) software was used for the construction of a transcriptional regulatory network.

Results

Differential gene expression in cancer tissue samples compared with normal samples. The gene expression profile of GSE 2514 was downloaded from the GEO database. Considering the scale of the calculations in the DCGL, pulmonary adenocarcinoma microarray data were analyzed and filtered using the limma and Affy packages to obtain the DEGs. A total of 1,379 DEGs were obtained with a P-value <0.05 and |logFC|>0.5. The DCGL in the R package was used to screen DCLs and DCGs. When Q<0.05 was used as the cut-off criterion, a total of 251 DCGs and 37,094 DCLs were obtained.

Identification of dysregulated pathways. In order to identify the dysregulated pathways in pulmonary adenocarcinoma samples, pathway enrichment analysis was performed using the online biological classification tool DAVID. A total of 10 pathways with a P-value <0.1 were enriched (Fig. 1). The most significant pathway was glycolysis/gluconeogenesis (P=5.49x10^{-6}). The other significant pathways included the cell cycle (P=8.84x10^{-5}), complement and coagulation cascades (P=0.018029) and DNA replication (P=0.044864). Ten significant pathways are listed in Table I.

TF regulatory network. All DCLs discovered in the present study were compared with the TRANSFAC database, and 10 TF-target gene pairs associated with pulmonary adenocarcinoma were observed, including eight TFs. The correlation difference (cor. diff.)-value (the sum of absolute values of the maximum absolute correlation for TF and target gene) indicated the degree of correlation between the TF and target gene (Table II). By mapping eight TFs to the TRANSFAC
database again, 366 regulation associations that were linked with those TFs were obtained. By processing these associations with Cytoscape software, eight TFs and 308 target genes were identified in the transcriptional regulatory network. Furthermore, several TFs and target genes belonging to DCGs did not appear among the DCLs in the present study, which indicated that those TFs may be synergistic with other factors. The results are illustrated in Fig. 2.

**Discussion**

The present study investigated gene expression profiles in patients with lung cancer and healthy controls to explore the biochemical pathways leading to the deterioration associated with pulmonary adenocarcinoma at the gene transcription level using a computational bioinformatics method. Furthermore, the study identified the most variable genes leading to pulmonary adenocarcinoma according to the associations between TFs and target genes. A total of 251 DCGs and 37,094 DCLs were obtained following DCGL analysis. These 251 DCGs were significantly enriched in 10 pathways. In addition, certain DCGs belonging to TF-target pairs did not appear among the DCLs. These TF-target gene pairs may have a synergistic effect with other TFs, which may provide novel insights for research into pulmonary adenocarcinoma.

In the present study, KEGG PATHWAY analysis was used to identify the dysregulated pathways in pulmonary adenocarcinoma samples; 10 pathways, including glycolysis/gluconeogenesis, the cell cycle and the complement and coagulation cascades, were highlighted. These pathways...
were associated with energy conversion, cell replication and the immune response. The most significant pathway was glycolysis/gluconeogenesis. Glycolysis involves the conversion of glucose into pyruvate and the generation of small amounts of adenosine triphosphate and nicotinamide adenine dinucleotide, and gluconeogenesis is a pathway leading to the synthesis of glucose from noncarbohydrate precursors. Gluconeogenesis is the reverse of glycolysis, but involves minor variations or alternative paths. A series of DCGs were included in the glycolysis/gluconeogenesis pathway, including phosphoglycerate kinase 1 (PGK1). Chen et al (23) suggested that four proteins, including the protein encoded by PGK1, that are involved in the glycolysis pathway are overexpressed in pulmonary adenocarcinoma; this overexpression was suggested to be associated with low survival rates in patients with pulmonary adenocarcinoma. A study on breast cancer demonstrated that tumors had abnormal bioenergetics, and patients with cancer showed a systematic loss of energy involving the interaction of tumor glycolysis and gluconeogenesis (24). The results of the differential co-expression analysis in the present study are in accordance with previous studies and indicated that the genes in the glycolysis/gluconeogenesis pathway have an important role in the development and progression of pulmonary adenocarcinoma.

A transcriptional regulatory network with 308 target genes and eight TFs was obtained using Cytoscape software. In this network, 19 target genes and all of the eight TFs [Fli-1 proto-oncoprotein, ETS transcription factor (FLI1), v-ets erythroblastosis
virus E26 oncogene homolog 2 (ETS2), hypoxia-inducible factor 1α (HIF1A), peroxisome proliferator-activated receptor γ (PPARG), lymphoid enhancer-binding factor 1 (LEFI), eti variant 4 (ETV4), jun D proto-oncogene (JUND) and vitamin D receptor (VDR) were DCGs. The most significant correlation of TF-target gene pairs was FLI1-cyclin-dependent kinase inhibitor 1A (CDKN1A) with a cor. diff.-value of 1.294162235. Other significant correlations included ETS2-urokinase-type plasminogen activator (PLAU) (cor. diff.-value, 1.225511357) and HIF1A-nuclear receptor subfamily 4, group A, member 1 (NR4A1) (cor. diff.-value, 1.222672148). Correlations of TF-target genes are listed in Table II.

FLI1 is a member of the ETS family of TFs characterized by the presence of the evolutionary conserved DNA-binding (ETS) domain, which recognizes the purine-rich GGA (A/T) core sequence (25). Sankar et al (26) suggested that the EWS RNA-binding protein 1 gene fuses with FLI1 to produce the EWS/FLI fusion protein, which is the abnormal TF that drives tumor development in Ewing sarcoma. CDKN1A is a type of CDK inhibitor (CKI), and is a cell cycle inhibitor gene regulated by VDR. The expression or stability of CKIs is reduced in tumors and leads to organ hyperplasia and increased tumor susceptibility (27). CDKN1A plays essential roles in the DNA damage response by inducing cell cycle arrest and directly inhibiting DNA replication, as well as by regulating fundamental processes, including apoptosis and transcription (28). The deletion of CDKN1A improves stem cell function and increases the lifespan of mice with dysfunctional telomerases without accelerating cancer formation (29). To date, genes whose expression has been reported to be repressed by EWS/FLI include CDKN1A (30,31).

It was also observed in the present study that DNA-binding transcription factors have three target genes. The expression of ETS2 has been shown to be elevated in certain cancer tissue samples and have a significant role in cancer progression (32). The genes regulated by ETS2 are those encoding enzymes that degrade the extracellular matrix, including stromelysin and collagen (33). The most significant relevant target gene to ETS2 is PLAU, which is currently used as a diagnostic marker (34). PLAU is closely correlated with the expression of a series of genes in lung cancer cell lines (35). However, in renal cancer cells, PLAU showed cancer cell-specific methylation that did not correlate well with expression status (36) and it was not specifically associated with colon cancer (37). The results of the cotransfection of the PLAU enhancer-CAT construct in the presence of increasing amounts of the ETS2-β-galactosidase expression vector showed that the co-expression of the ETS dominant-negative protein resulted in the almost complete inhibition of the PLAU enhancer induction (38). The second relevant gene to ETS2 is JUNB, which represents an important target in diseases associated with cancer and fibrosis (39). The expression of JUNB is inactivated by methylation in chronic myeloid leukemia (40). The third relevant gene to ETS2 is matrix metalloproteinase-9 (MMP9). The expression of MMP9 in non-small cell lung cancer contributes to tumor cell invasiveness (41). ETS can regulate MMPs, and MMP9 expression was shown to be suppressed by ETS blockage through overexpression of a dominant-negative form of ETSI (42).

HIF1A encodes a pivotal TF that regulates angiogenesis by inducing the expression of vascular endothelial growth factor, interleukin-8 and a basic fibroblast growth factor (43). HIF1A is considered to be one of the key regulators of tumor angiogenesis. NR4A1 is the target gene of HIF1A. NR4A1 was demonstrated to reduce the migration of normal and breast cancer cell lines (44). Results regarding the knockdown or overexpression of NR4A1 in lung cancer cells suggest that this receptor exhibits pro-oncogenic activity and enhances cell survival and/or proliferation (45).

The other TFs examined in the present study showed less correlation with their target genes, but these also had an important role in the molecular mechanism of the disease. PPARG is a member of the PPAR family, which has a pivotal role in adipogenesis and glucose homeostasis (46). LEF1 mRNA levels are important statistical metrics in cancer. A study on ovarian, fallopian tube and peritoneal cancer indicated that LEF1 overexpression may be predictive of poor overall survival (47). Furthermore, ETV4 was demonstrated to be a useful marker in study on prostate cancer (48), and JUND and its target gene PLAU were shown to have an important role in gene expression in cancer cells (49). The TF with the least correlation was VDR, which only had five target genes on the map. This indicates that VDR is an important site for carcinogenesis.

In conclusion, a total of 10 pathways were enriched in pulmonary adenocarcinoma tissue samples, and the most significant pathway of DCGs in the present study was glycolysis/gluconeogenesis. Ten TF-target gene pairs that were associated with pulmonary adenocarcinoma were discovered in the transcriptional regulatory network. Co-expression analysis of these TFs and target genes may provide novel ideas for cancer research, and the results of the present study may provide groundwork enabling the investigation of the most variable genes leading to the development and progression of pulmonary adenocarcinoma. However, further experiments are required to confirm these observations.

References


