RESEARCH COMMUNICATION

Bioinformatics Analysis Reveals Connection of Squamous Cell Carcinoma and Adenocarcinoma of the Lung

Wei-Dong Fan, Xian-Quan Zhang*, Hui-Lin Guo, Wei-Wei Zeng, Ni Zhang, Qian-Qian Wan, Wen-Yao Xie, Jin Cao, Chang-Hua Xu

Abstract

    Squamous cell carcinoma and adenocarcinoma are the major histological types of non-small cell lung cancer. Because they differ on the basis of histopathological and clinical characteristics and their relationship with smoking, their etiologies may be different; for example, different tumor suppressor genes may be related to the genesis of each type. We used microarray data to construct three regulatory networks to identify potential genes related to lung adenocarcinoma and squamous cell carcinoma and investigated the similarity and specificity of them. In the network, some of the observed transcription factors and target genes had been previously proven to be related to lung adenocarcinoma and squamous cell carcinoma. We also found some new transcription factors and target genes related to SCC. The results demonstrated that regulatory network analysis is useful in connection analysis between lung adenocarcinoma and squamous cell carcinoma.

Keywords: Lung cancer - squamous cell carcinoma - adenocarcinoma - regulatory network

Introduction

    Lung cancer is the second most common type of cancer in the world and is the most common cause of cancer-related death in men and women; it is responsible for 1.3 million deaths annually worldwide (Yildiz et al., 2011). Squamous cell carcinoma and adenocarcinoma are the major histological types of non-small cell lung cancer (NSCLC). Because they differ on the basis of histopathological and clinical characteristics and their relationship with smoking, their etiologies maybe different; for example, a different tumor suppressor gene may be related to the genesis of each type. Accounting for 25% of lung cancers (Travis, 2002), squamous cell lung carcinoma usually starts near a central bronchus. A hollow cavity and associated necrosis are commonly found at the center of the tumor. Adenocarcinoma accounts for 40% of non-small-cell lung cancers (Travis, 2002). It usually originates in peripheral lung tissue. Most cases of adenocarcinoma are associated with smoking; however, among people who have never smoked ("never-smokers"), adenocarcinoma is the most common form of lung cancer (Subramanian and Govindan, 2007).

    Many genetic alterations associated with lung cancer have been reported to date. Changes in oncogenes including activation of genes in the myc and ras families (Johnson et al., 1987; Rodenhuis et al., 1987; Gemma et al., 1988; Rodenhuis et al., 1988) and inactivation of tumor suppressor genes including several tumor suppressor genes

(TP53, CDKN2A and STK11, NF1, ATM, RB1, APC, etc) along with tyrosine kinase genes (KRAS, EGFR, ERBB, EPHA3, NRAS, KDR, FGFR4 and NTRK etc) that may function as proto-onco-genes. Further insight into the gene alterations underlying lung cancer was involved in signaling pathway. Additionally, there is a significant excess of mutations and copy number alterations in genes from the MAPK, p53, Wnt cell cycle and mTOR signaling pathways, suggesting links to the disease (Ding et al., 2008).

    Microarray gene expression profiling has recently been used to define prognostic signatures in patients with lung carcinomas (Garber et al., 2001; Miura et al., 2002; Beer et al., 2002; Ramaswamy et al., 2003); A high-throughput microarray experiment was designed to analyze genetic expression patterns and identify potential genes to target for AC (Li et al., 2006). However, whereas several studies have investigated gene expression profiles in lung AC, there have been no similar large studies that have investigated in SCC. AC and SCC are closely related in humans, especially among smokers. In order to investigate the similarity and specificity of SCC and ACC in gene expression profiles, we found differentially expressed genes between them and constructed three regulatory networks, the common regulatory network, the AC specific network and the SCC specific network. Furthermore, a set of transcription factors which regulate the differently expression genes play an important role in the progress of SCC and AC were identified in our network.

*Department of Oncology, the Second Affiliated Hospital of Chongqing Medical University, Chongqing, China *For correspondence: zhangxianquan123@gmail.com*

Materials and Methods

Affymetrix microarray data

Our data was obtained from Bhattacharjee et al. (2001), which are based on the Affymetrix human U95A oligonucleotide arrays. Only 161 chips are available. The 161 specimens include squamous cell lung carcinomas (n=21), adenocarcinoma (n=123) and normal lung (n=17) specimens. Total 8742 genes were detected in our experiment.

Pathway data

Kyoto Encyclopedia of Genes and Genomes (KEGG) is a collection of online databases dealing with genomes, enzymatic pathways, and biological chemicals (Kanehisa, 2002). The 'pathway' database records networks of molecular interactions in the cells, and variants of them specific to particular organisms (http://www.genome.jp/kegg/).

Regulation data

There are approximately 2600 proteins in the human genome that contain DNA-binding domains, and most of these are presumed to function as transcription factors (Wachi et al., 2005). The combinatorial use of a subset of the approximately 2000 human transcription factors easily accounts for the unique regulation of each gene in the human genome during development (Brivanlou and Darnell, 2002).

These transcription factors are grouped into 5 super class families, based on the presence of conserved DNA-binding domains. TRANSFAC database contains data on transcription factors, and their experimentally-proven binding sites and regulated genes (Wingender, 2008).

Transcriptional Regulatory Element Database (TRED) has been built to increase the needs of an integrated repository for both cis- and trans- regulatory elements in mammals (Jiang et al., 2007). TRED did the curation for transcriptional regulation information, including transcription factor binding motifs and experimental evidence. The curation is currently focusing on target genes of 36 cancer-related TF families.

Combined the two regulation datasets, total 7201 regulatory relationships between 375 TFs and 2650 target genes were collected.

Differentially expressed genes (DEGs) analysis

The t-test method was used to identify DEGs. To circumvent the multi-test problem which might induce too much false positive results, the BH method (Fromm et al., 1986) was used to adjust the raw P-values into false discovery rate (FDR). The original expression datasets from all conditions were processed into expression estimates using the RMA and quantile method with the default settings implemented by Affy package in R, and then constructed the linear model. The DEGs only with the fold change value larger than 1.2 and FDR less than 0.05 were selected.

Pathway enrichment analysis

In order to facilitate the functional annotation and analysis of large lists of genes in our network, DAVID (Huang da et al., 2009) (The Database for Annotation, Visualization and Integrated Discovery) bioinformatics resources were used. DAVID consists of an integrated biological knowledgebase and analytic tools aimed at systematically extracting biological meaning from large gene/protein lists.

Regulation network construction

Using the regulation data that have been collected from TRANSFAC database and TRED database, we matched the relationships between differentially expressed TFs and their differentially expressed target genes. We built the regulatory networks by Cytoscape (Shannon et al., 2003). We got 91 regulatory relationships between 21 TFs and their 72 differentially expressed target genes in adenocarcinoma and 314 regulatory relationships between 39 TFs and their 204 differentially expressed target genes in squamous cell lung carcinomas. By integrating the two networks, fifty-seven overlapping regulatory relationships between 11 TFs and their 50 target genes were obtained.

Results

Differentially expressed genes (DEGs) selection

After microarray analysis, the differentially expressed genes with the fold change value larger than 1.2 and FDR less than 0.05 were selected. We found 1376 DEGs between normal lung specimens and adenocarcinoma, and 2292 DEGs between normal lung specimens and squamous cell lung carcinomas. One thousand and fifty overlapping DEGs were found between the 1376 DEGs and 2292 DEGs, wherein 1039 DEGs changes in the same direction. PHF3, HAS2, STX1A, PRPF4, NCAPD2, CEL, GPSN2, NMU, LRP10 and APOBEC3F expressed at a low level in adenocarcinoma, but a high level in squamous cell lung carcinomas. Only GABBR2 has high expression in adenocarcinoma while low expression in squamous cell lung carcinomas. The expression values of PHF3, HAS2, STX1A, PRPF4, NCAPD2, and GABBR2 in three specimens were demonstrated in Figure 1. The different expression values suggesting that these TFs may play an important role in lung carcinoma. There are 326 specific genes only found in AC and 1242 specific genes only found in SCC, which suggesting that although the
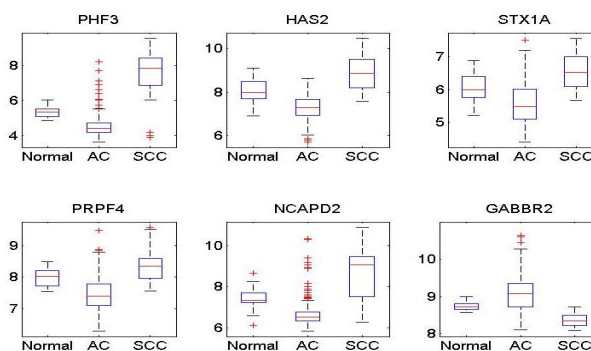

Figure 1. Expression Value of Part of DEGs Between AC and SCC in Three Specimens. The ordinate is gene expression value, 'Normal' represents normal lung specimens, 'AC' represents adenocarcinoma and 'SCC' represents squamous cell lung carcinomas

**Table 1. GO Enrichment Analysis (only list top 10 of enriched GO terms)**

Class	GO ID	GO Term	P value	FDR
SCC special	GO:0008283	cell proliferation	0	0
SCC special	GO:0010033	response to organic substance	5.85E-14	3.54E-11
SCC special	GO:0051239	regulation of multicellular organismal process	9.71E-13	3.76E-10
SCC special	GO:0010628	positive regulation of gene expression	6.49E-12	2.10E-09
SCC special	GO:0009888	tissue development	7.73E-12	2.34E-09
SCC special	GO:0060548	negative regulation of cell death	1.78E-11	5.23E-09
SCC special	GO:0009611	response to wounding	3.91E-11	1.05E-08
SCC special	GO:0009891	positive regulation of biosynthetic process	1.50E-10	3.45E-08
SCC special	GO:0009605	response to external stimulus	2.97E-10	5.98E-08
SCC special	GO:0048468	cell development	5.51E-10	9.36E-08
AC special	GO:0045944	positive regulation of transcription from RNA polymerase II promoter	4.87E-12	9.11E-09
AC special	GO:0001779	natural killer cell differentiation	1.01E-07	4.43E-05
AC special	GO:0030217	T cell differentiation	3.14E-06	0.000661
AC special	GO:0046544	development of secondary male sexual characteristics	4.83E-06	0.000954
AC special	GO:0007565	female pregnancy	5.85E-06	0.001132
AC special	GO:0030183	B cell differentiation	5.97E-06	0.001134
AC special	GO:0030218	erythrocyte differentiation	7.43E-06	0.001383
AC special	GO:0001889	liver development	1.18E-05	0.002008
AC special	GO:0046449	creatinine metabolic process	1.45E-05	0.002295
AC special	GO:0045931	positive regulation of mitotic cell cycle	1.98E-05	0.002949
common	GO:0065008	regulation of biological quality	1.18E-10	1.15E-06
common	GO:0032583	regulation of gene-specific transcription	3.67E-08	5.08E-05
common	GO:0010604	positive regulation of macromolecule metabolic process	4.81E-08	5.82E-05
common	GO:0060395	SMAD protein signal transduction	2.56E-07	0.000133
common	GO:0007498	mesoderm development	4.28E-07	0.000186
common	GO:0006917	induction of apoptosis	1.04E-06	0.000282
common	GO:0030518	steroid hormone receptor signaling pathway	8.34E-07	0.000282
common	GO:0042448	progesterone metabolic process	1.06E-06	0.000282
common	GO:0045597	positive regulation of cell differentiation	2.85E-06	0.000601
common	GO:0051704	multi-organism process	3.82E-06	0.000711
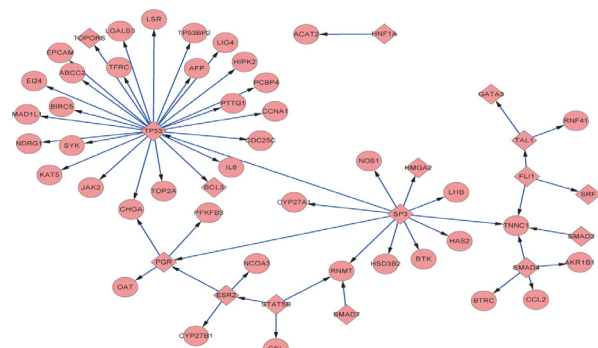
Figure 2. The Common Regulatory Network Between AC and SCC. Same subnets were found in this common regulatory network. Rhombus (◊) represents transcription factors and circle (O) represents target genes

pathogenesis of AC and SCC were significantly similar, they are also have their own specificity.

Regulation network construction and analysis

Because AC and SCC are two subtypes of non-small cell lung cancer, their regulatory networks may have the same subnets. We obtained microarray data from Bhattacharjee et al (2001). After microarray analysis, the differentially expressed genes with the fold change value larger than 1.2 and FDR less than 0.05 were selected. We got 91 regulatory relationships between 21 TFs and their 72 differently expressed target genes in adenocarcinoma and 314 regulatory relationships between 39 TFs and their 204 differentially expressed target genes in squamous cell
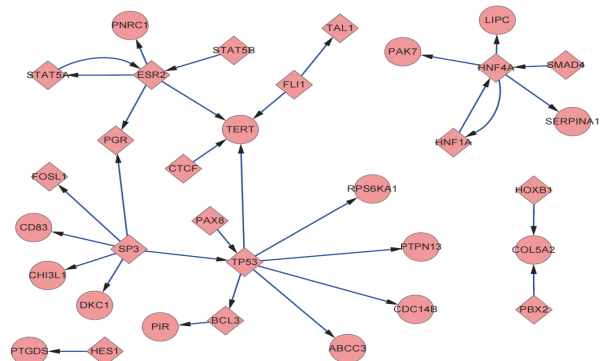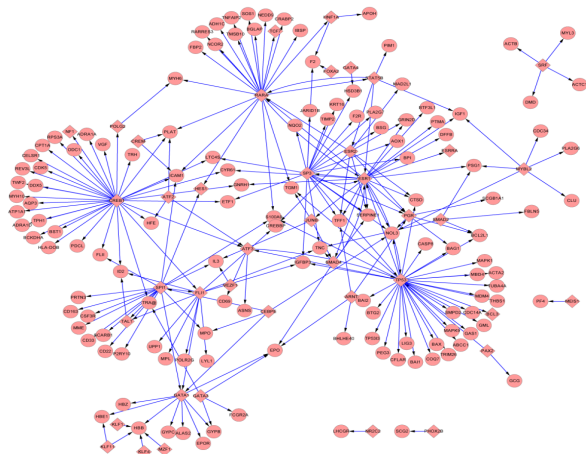

Figure 3. The AC specific regulatory network. Rhombus (◊) represents transcription factors and circle (O) represents target genes

lung carcinomas. There are 57 overlapping regulatory relationships between AC and SCC. By integrating the regulatory relationships above, three regulatory networks (Figure 2-4) were built between TFs and their target genes. As we can see in Figure 2, TP53, SP3, PGR, SP3, SMAD4 and ESR2 with higher degrees form a local network which suggesting that these TFs may play an important role both in AC and SCC. Of these TFs, TP53 and PGR have high expression in the two subtypes while SP3, SMAD4 and ESR2 have low expression.

There is a specific regulatory network in AC (Figure 3). Compared to the common regulatory network, the AC specific network adds some relationships related with TP53, ESR2, and SP3 etc, and the regulatory relationships of HNF4A only found in AC specific network, which

Table 2. DAVID Enrichment Analysis (only list enriched pathways with the FDR < 0.1)

Class	Term	Description	P-Value	FDR
SCC Special	hsa04115	p53 signaling pathway	4.10E-06	5.13E-04
SCC Special	hsa05200	Pathways in cancer	3.60E-04	0.022238
SCC Special	hsa05221	Acute myeloid leukemia	7.10E-04	0.029145
SCC Special	hsa05210	Colorectal cancer	9.50E-04	0.029253
SCC Special	hsa05416	Viral myocarditis	0.002062338	0.050302
SCC Special	hsa04630	Jak-STAT signaling pathway	0.002421626	0.049257
SCC Special	hsa04640	Hematopoietic cell lineage	0.005398849	0.092144
SCC Special	hsa04210	Apoptosis	0.005712095	0.085618
SCC Special	hsa05215	Prostate cancer	0.006378245	0.085036
AC Special	hsa05220	Chronic myeloid leukemia	0.001806333	0.074797
AC Special	hsa04950	Maturity onset diabetes of the young	0.003016788	0.062894
common	hsa04110	Cell cycle	1.50E-04	0.011012

Figure 4. The SCC specific regulatory network.
Rhombus (◊) represents transcription factors and circle (O) represents target genes

suggesting that the regulatory relationships caused by HNF4A play an important role in carcinogenesis of AC.

The SCC specific regulatory network was built as shown in figure 4. In this network, there are also some new regulatory relationships related with TP53 and SP3. The most significant difference is that a large number of regulatory relationships related with CREB1, ESR1, RARA, SPI1 were found in this network and these TFs does not appeared in AC specific network.

GO analysis of the regulation network

To investigate the function differences of AC and SCC network, several Gene Ontology (GO) categories were enriched among these genes in the common regulatory network, AC specific regulatory network and SCC specific regulatory network. The gene function of the common regulatory network contains regulation of biological quality, regulation of gene-specific transcription, positive regulation of macromolecule metabolic process and so on. The gene function of AC specific regulatory network contains positive regulation of transcription from RNA polymerase II promoter, natural killer cell differentiation, T cell differentiation and so on. The gene function of SCC specific regulatory network contains cell proliferation, response to organic substance, regulation of multicellular organismal process and so on (Table 1).

DAVID analysis of the regulation network

We used DAVID to perform pathway enrichment of the

three regulatory networks. Seven KEGG pathways were enriched among these genes in the common regulatory network, including 'Cell cycle', 'Wnt signaling pathway' and so on. For AC specific regulatory network, pathways including 'Cell cycle', 'Wnt signaling pathway', 'ErbB signaling pathway' etc were enriched; for the SCC specific regulatory network, the enriched pathways contains 'p53 signaling pathway', 'TGF-beta signaling pathway', 'Focal adhesion', 'Jak-STAT signaling pathway' and so on (Table 2).

Discussion

The recent development of cDNA microarray or cDNA chip technology, a high-throughput method of monitoring gene expression, has made it possible to analyze the expression of thousands of genes at once (Ermolaeva et al., 1998; Joseph, 1996; Khan et al., 1998; Schena et al., 1995). Consequently, regulatory networks of TFs and their target genes can now be constructed on the basis of the altered expression of multiple genes in different tumor specimens. In the study reported here, we analyzed the differentially expressed genes of AC and SCC, and constructed three regulatory networks, the common regulatory network, the AC specific network and the SCC specific network. Furthermore, gene ontology analysis and DAVID analysis were performed to analyse the functional annotation of genes and involved pathways in the three networks.

From the result of common regulation network construction between AC and SCC, we could find that TP53, SP3, PGR, SMAD4 and ESR2 are hub nodes in our common regulatory network and some of them were proved to be related with lung carcinomas by previous studies (Beattie et al., 1985 Sekido et al., 1998; Toyooka et al., 2003; Zhu et al., 2003; Marquez-Garban et al., 2007). Of these TFs, TP53 and PGR have high expression in the two subtypes while SP3, SMAD4 and ESR2 have low expression, which suggesting that these TFs may play an important role both in AC and SCC. These expression profiles also found in previous studies (Takahashi et al., 1989; Bodner et al., 1992; Kosaka et al., 2009) of lung cancer.

TP53 (also known as p53), is a tumor suppressor gene that encodes tumor suppressor protein. Tumor protein p53 is a DNA-binding protein which responds to diverse cellular stresses to regulate target genes that induce cell cycle arrest, apoptosis, senescence, DNA

repair, or changes in metabolism. TP53 was identified to mutate in lung carcinoma (Le Calvez et al., 2005). Furthermore, the frequency of TP53 mutations in invasive tumors was significantly higher than those in noninvasive ones (Nakanishi et al., 2009). SP3 belongs to a family of 16 different transcription factors containing a highly conserved zinc finger DNA-binding domain (Philipsen and Suske, 1999). SP3 is over-expressed in a variety of cancers including gastric, colorectal, pancreatic, epidermal, thyroid, breast and lung cancers. Previous studies (Colon et al., 2011) showed that Sp proteins mediate a number of genes involved in cell proliferation, survival and angiogenesis.

HNF4A is a TF which only observed in the AC specific network; this suggested that the regulatory relationships caused by HNF4A may play an important role in carcinogenesis of AC. HNF4A is a nuclear transcription factor that binds DNA as a homodimer. The encoded protein controls the expression of several genes, including hepatocyte nuclear factor 1 alpha, a transcription factor which regulates the expression of several hepatic genes. Although the role of HNF4A has not been confirmed to be related with lung carcinoma, some evidence suggests that the average expression levels of HNF4a gene in some lung disease were modestly increased compared to normal human lungs. In comparison with normal human lungs, the averages of HNF4A gene expression levels were significantly higher in human adenocarcinoma. In contrast to adenocarcinoma, the average expression levels of HNF4A in squamous cell carcinomas were significantly reduced (Qu et al., 2009).

A large number of regulatory relationships related with CREB1, ESR1, RARA, SPI1 were found in SCC specific network and these TFs does not appeared in AC specific network. There are 40 differentially expressed target genes of CREB1, but only 6 of them were found in AC specific network, most of them regulate genes in SCC specific network. There are 24 differentially expressed target genes of RARA, but only 4 of them were found in AC specific network, most of them regulate genes in SCC specific network. We proposed a hypothesis that the high expression profile of CREB1, ESR1, RARA and SPI1 and their corresponding regulatory relationship are of importance in the carcinogenesis of SCC.

In the past decade, many signaling pathways are identified associated with cancer development, such as JAK/STAT, MAPK/ERK, PI3K/AKT, NF-kB, Wnt, TGF-β, etc (Dreesen and Brivanlou, 2007). Identically, these pathways play a significant role in lung carcinomas. Mutations in the BRAF and KRAS genes occur in non-small-cell lung cancer patients, which induces the activation of extracellular signal-regulated kinase (ERK)-1/2 (MAPK) pathway and the development of lung carcinoma (Ji et al., 2007). The Wnt signaling pathway, originally identified in Drosphila melanogaster, is highly conserved in evolutionary pathways involved in homeostasis, cell proliferation, differentiation, motility, and apoptosis. It is deregulated in a number of cancers, including lung cancer. TGF (Transforming growth factor β) - beta signaling pathway were proved to be related with lung carcinoma. Cell-autonomous TGF-β

signaling is required for both induction and maintenance of in vitro invasiveness and metastasis during late-stage tumorigenesis (Martin, 1998).

The basic understanding of the mechanisms underlying the function of AC and SCC genes is important. A deeper understanding of transcription factors and their regulated genes remain an area of intense research activity in futures. Our regulatory network is useful in investigating the connection between AC and SCC. We also find some new transcription factors and target genes related to AC and SCC; besides, many pathways such as pathways in cancer, p53 signaling pathway and Jak-STAT signaling pathway have been linked with SCC by our method. However, further experiments are still needed to confirm the conclusion.

References

Beattie CW, Hansen NW, Thomas PA (1985). Steroid receptors in human lung cancer. *Cancer Res*, **45**, 4206-14.

Beer DG, Kardia SL, Huang CC, et al (2002). Gene-expression profiles predict survival of patients with lung adenocarcinoma. *Nat Med*, **8**, 816-24.

Bhattacharjee A, Richards WG, Staunton J, et al (2001). Classification of human lung carcinomas by mRNA expression profiling reveals distinct adenocarcinoma subclasses. *Proc Natl Acad Sci U S A*, **98**, 13790-5.

Bodner SM, Minna JD, Jensen SM, et al (1992). Expression of mutant p53 proteins in lung cancer correlates with the class of p53 gene mutation. *Oncogene*, **7**, 743-9.

Brivanlou AH, Darnell JE, Jr (2002). Signal transduction and the control of gene expression. *Science*, **295**, 813-8.

Colon J, Basha MR, Madero-Visbal R, et al (2011). Tolfenamic acid decreases c-Met expression through Sp proteins degradation and inhibits lung cancer cells growth and tumor formation in orthotopic mice. *Invest New Drugs*, **29**, 41-51.

Ding L, Getz G, Wheeler DA, et al (2008). Somatic mutations affect key pathways in lung adenocarcinoma. *Nature*, **455**, 1069-75.

Dreesen O, Brivanlou AH (2007). Signaling pathways in cancer and embryonic stem cells. *Stem Cell Rev*, **3**, 7-17.

Ermolaeva O, Rastogi M, Pruitt KD, et al (1998). Data management and analysis for gene expression arrays. *Nat Genet*, **20**, 19-23.

Fromm M, Littman P, Raney RB, et al (1986). Late effects after treatment of twenty children with soft tissue sarcomas of the head and neck. Experience at a single institution with a review of the literature. *Cancer*, **57**, 2070-6.

Garber ME, Troyanskaya OG, Schluens K, et al (2001). Diversity of gene expression in adenocarcinoma of the lung. *Proc Natl Acad Sci U S A*, **98**, 13784-9.

Gemma A, Nakajima T, Shiraishi M, et al (1988). myc family gene abnormality in lung cancers and its relation to xenotransplantability. *Cancer Res*, **48**, 6025-8.

Huang da W, Sherman BT, Lempicki RA (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc*, **4**, 44-57.

Ji H, Wang Z, Perera SA, et al (2007). Mutations in BRAF and KRAS converge on activation of the mitogen-activated protein kinase pathway in lung cancer mouse models. *Cancer Res*, **67**, 4933-9.

Jiang C, Xuan Z, Zhao F, Zhang MQ (2007). TRED: a transcriptional regulatory element database, new entries and other development. *Nucleic Acids Res*, **35**, D137-40

Johnson BE, Ihde DC, Makuch RW, et al (1987). myc family

oncogene amplification in tumor cell lines established from small cell lung cancer patients and its relationship to clinical status and course. *J Clin Invest*, **79**, 1629-34.

Joseph D, Lolita P, Patrick OB, et al (1996). Use of a cDNA microarry to analyse gene expression patterns in human cancer. *Nature Genetics*, **14**, 457-60.

Kanehisa M (2002). The KEGG database. *Novartis Found Symp*, **247**, 91-101 (discussion 101-3, 119-28, 244-52).

Khan J, Simon R, Bittner M, et al (1998). Gene expression profiling of alveolar rhabdomyosarcoma with cDNA microarrays. *Cancer Res*, **58**, 5009-13.

Kosaka T, Yatabe Y, Onozato R, et al (2009). Prognostic implication of EGFR, KRAS, and TP53 gene mutations in a large cohort of Japanese patients with surgically treated lung adenocarcinoma. *J Thorac Oncol*, **4**, 22-9.

Le Calvez F, Mukeria A, Hunt JD, et al (2005). TP53 and KRAS mutation load and types in lung cancers in relation to tobacco smoke: distinct patterns in never, former, and current smokers. *Cancer Res*, **65**, 5076-83.

Li R, Wang H, Bekele BN, et al (2006). Identification of putative oncogenes in lung adenocarcinoma by a comprehensive functional genomic approach. *Oncogene*, **25**, 2628-35.

Marquez-Garban DC, Chen HW, Fishbein MC, et al (2007). Estrogen receptor signaling pathways in human non-small cell lung cancer. *Steroids*, **72**, 135-43.

Martin O, Karl-heinz H, Hartmut B (1998). TGF-beta signaling is necessary for carcinoma cell invasiveness and metastasis. *Current Biology,* **8**, 1243-52.

Miura K, Bowman ED, Simon R, et al (2002). Laser capture microdissection and microarray expression analysis of lung adenocarcinoma reveals tobacco smoking- and prognosis-related molecular profiles. *Cancer Res*, **62**, 3244-50.

Nakanishi H, Matsumoto S, Iwakawa R, et al (2009). Whole genome comparison of allelic imbalance between noninvasive and invasive small-sized lung adenocarcinomas. *Cancer Res*, **69**, 1615-23.

Philipsen S, Suske G (1999). A tale of three fingers: the family of mammalian Sp/XKLF transcription factors. *Nucleic Acids Res*, **27**, 2991-3000.

Qu P, Roberts J, Li Y, et al (2009). Stat3 downstream genes serve as biomarkers in human lung carcinomas and chronic obstructive pulmonary disease. *Lung Cancer*, **63**, 341-7.

Ramaswamy S, Ross KN, Lander ES and Golub TR, (2003). A molecular signature of metastasis in primary solid tumors. *Nat Genet*, **33**, 49-54.

Rodenhuis S, Slebos RJ, Boot AJ, et al (1988). Incidence and possible clinical significance of K-ras oncogene activation in adenocarcinoma of the human lung. *Cancer Res*, **48**, 5738-41.

Rodenhuis S, van de Wetering ML, Mooi WJ, et al (1987). Mutational activation of the K-ras oncogene. A possible pathogenetic factor in adenocarcinoma of the lung. *N Engl J Med*, **317**, 929-35.

Schena M, Shalon D, Davis RW, Brown PO (1995). Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science*, **270**, 467-70.

Sekido Y, Fong KM, Minna JD (1998). Progress in understanding the molecular pathogenesis of human lung cancer. *Biochim Biophys Acta*, **1378**, F21-59.

Shannon P, Markiel A, Ozier O, et al (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res*, **13**, 2498-504.

Subramanian J, Govindan R (2007). Lung cancer in never smokers: a review. *J Clin Oncol*, **25**, 561-70.

Takahashi T, Nau MM, Chiba I, et al (1989). p53: a frequent target for genetic abnormalities in lung cancer. *Science*, **246**, 491-4.

Toyooka S, Tsuda T, Gazdar AF (2003). The TP53 gene, tobacco exposure, and lung cancer. *Hum Mutat*, **21**, 229-39.

Travis WD (2002). Pathology of lung cancer. *Clin Chest Med*, **23**, 65-81.

Wachi S, Yoneda K, Wu R (2005).I nteractome-transcriptome analysis reveals the high centrality of genes differentially expressed in lung cancer tissues. Bioinformatics, 21, 4205-8.

Wingender E (2008). The TRANSFAC project as an example of framework technology that supports the analysis of genomic regulation. *Brief Bioinformatics*, **9**, 326-32.

Yildiz O, Buyuktas D, Ekiz E, et al (2011). Facial nerve palsy: an unusual presenting feature of small cell lung cancer. *Case Rep Oncol*, **4**, 35-8.

Zhu WG, Srinivasan K, Dai Z, et al (2003). Methylation of adjacent CpG sites affects Sp1/Sp3 binding and activity in the p21(Cip1) promoter. *Mol Cell Biol*, **23**, 4056-65.