Identifying differentially expressed genes and pathways in two types of non-small cell lung cancer: adenocarcinoma and squamous cell carcinoma

J. Liu, X.Y. Yang and W.J. Shi

Department of Thoracic Surgery, Shengjing Hospital,
China Medical University, Heping District, Shenyang, Liaoning, China

Corresponding author: W.J. Shi
E-mail: shijunwensjw@hotmail.com

ABSTRACT. Non-small cell lung carcinoma, NSCLC, accounts for 80-85% of lung cancers. NSCLC can be mainly divided into two types: adenocarcinoma (ADC) and squamous cell carcinoma (SCC). The purpose of our study was to identify and differentiate the pathogenesis of ADC and SCC at the molecular level. The gene expression profiles of ADC and SCC were downloaded from Gene Expression Omnibus under accession No. GSE10245. Accordingly, differentially expressed genes (DEGs) were identified by the limma package in R language. In addition, DEGs were functionally analyzed by Gene Ontology and Kyoto Encyclopedia of Genes and Genomes pathway enrichment. A total of 4124 DEGs were identified, including CDK1, CDK2, CDK4, and SKP2. The DEGs were mainly involved in 16 pathways related to cell proliferation, cell signal transduction and metabolism. We conclude that the molecular mechanisms of ADC and SCC are considerably different, and that they are involved in immune response, cell signal transduction, metabolism, cell division, and cell proliferation. Therefore, the two diseases should be

treated differently. This study offers new insight into the diagnosis and therapy of these two types of lung cancer.

Key words: Non-small cell lung cancer; Adenocarcinoma; Squamous cell carcinoma; Differentially expressed genes

INTRODUCTION

Lung cancer is the most frequent cause of cancer deaths in the North America and Europe (Ferlay et al., 2007). Lung cancers are classified into four major sub-types according to their histopathological presentation: small cell lung cancer, squamous cell carcinoma (SCC), adenocarcinoma (ADC), and large cell carcinoma. The latter three are all included in the non-small cell lung cancer (NSCLC) classification. ADC and SCC collectively account for almost 70% of lung cancers (SCC 40%, ADC 30%) and are known to be the most important subtypes of NSCLC (ATS and ERS, 1997).

On the basis of the classification of histopathological features, different methods are designed for the treatment of various types of lung cancers. However, several researchers have found discrepancies where the same type of tumor classified by histopathological features shows dramatically different responses to the same therapy (Hou et al., 2010). Therefore, more accurate classification methods are urgently needed.

Recently, several studies have attempted to distinguish non-small cell lung ADC from SCC at the molecular level (Daraselia et al., 2012). The signatures of ADC and SCC differ in the composition of genes. For ADC, genes are more correlated with tight junction and cell adhesion molecules, and on the contrary, the SCC signature favors the genes involved in cell communication (Hou et al., 2010). Also, Terry et al. (2010) report that p63 is the single best marker to separate ADC from SCC, with the sensitivity of 84% and the specificity of 85%. Further analysis identifies p63, TTF1, CK5/6, and Napsin A as the optimal panel to isolate ADC from SCC (Kim et al., 2013). Novel biomarkers (miR-93, miR-205, miR-221, and let-7e) for differential diagnosis and prognosis of NSCLC subtypes are also found by means of identifying differential expression profiles of miRNAs in ADC and SCC (Zhang et al., 2012). The improved classification of NSCLC is meaningful to clinical therapy.

In this study, we downloaded the gene expression profiles of SCC and ADC to identify the differentially expressed genes (DEGs) between SCC and ADC with bioinformatic methods. Furthermore, we used Gene Ontology (GO) enrichment analysis and biological pathway enrichment analysis to annotate the function of these DEGs. All the information obtained will hence advance the diagnosis of SCC and ADC.

MATERIAL AND METHODS

Gene expression profile of ADC and SCC

The gene expression profile of GSE10245 (Kuner et al., 2009) was downloaded from the public functional genomics data repository GEO (Gene Expression Omnibus, http://www.ncbi.nlm.nih.gov/geo/) database. A total of 58 chips, including 40 chips of ADC samples and 18 chips of SCC samples, were available based on the GPL570 platform.

DEG analysis

The original microarray data were divided into two groups, ADC and SCC. We analyzed the derived genetic data with the limma package in R language (v.2.13.0) (R Development Core Team, 2008). Microarray data were normalized with, Robust Multichip Averaging (Irizarry et al., 2003). The limma (Smyth, 2004) package, as a linear regression model, was applied to compare the ADC and SCC samples. The Benjamini-Hochberg method was then used to adjust the P values. A false discovery rate (FDR) less than 0.05 was chosen as the cut-off criterion.

GO enrichment analysis

GO project (The Gene Ontology Consortium, 2008; http://www.geneontology.org/) provides for community use in annotating genes, gene products and sequences (Harris et al., 2008), and GO (Scagliotti et al., 2008) is used for classifying gene function and location. The cluster Profiler (Yu et al., 2012) package was applied to perform GO enrichment analysis of cellular composition, molecular function and biological process.

Biological pathway enrichment analysis

To explore the differences of ADC and SCC at the molecular level, we focused on the biological pathway. Kyoto Encyclopedia of Genes and Genomes (KEGG) is an online database for systematic analysis of gene functions (Kanehisa and Goto, 2000). The entire metabolic and non-metabolic pathways were downloaded from KEGG pathway database (Irizarry et al., 2003), and cluster analysis with Database for Annotation, Visualization and Integrated Discovery (DAVID) was conducted. A count number larger than 2 and P value less than 0.05 were used as cut-off criteria.

RESULTS

Identification of DEGs between ADC and SCC

For dataset GSE10245, a total of 6562 probes were found to be differentially expressed between ADC and SCC samples at an FDR of 0.05, which corresponded to 4124 DEGs.

Functional annotation of DEGs

To investigate the function of DEGs between ADC and SCC, the DEGs were mapped to the GO database. The cellular compartments, in which the DEGs were most located, such as condensed chromosome, centrioles and kinetochore, are shown in Figure 1. Similarly, the molecular functions of DEGs, for instance, DNA binding, GTPase activator activity, kinase C inhibitor activity and some metabolites, are displayed in Figure 2. Moreover, Figure 3 shows the biological processes of DEGs, which were associated with DNA replication and mitosis linked to cell division and cell proliferation.
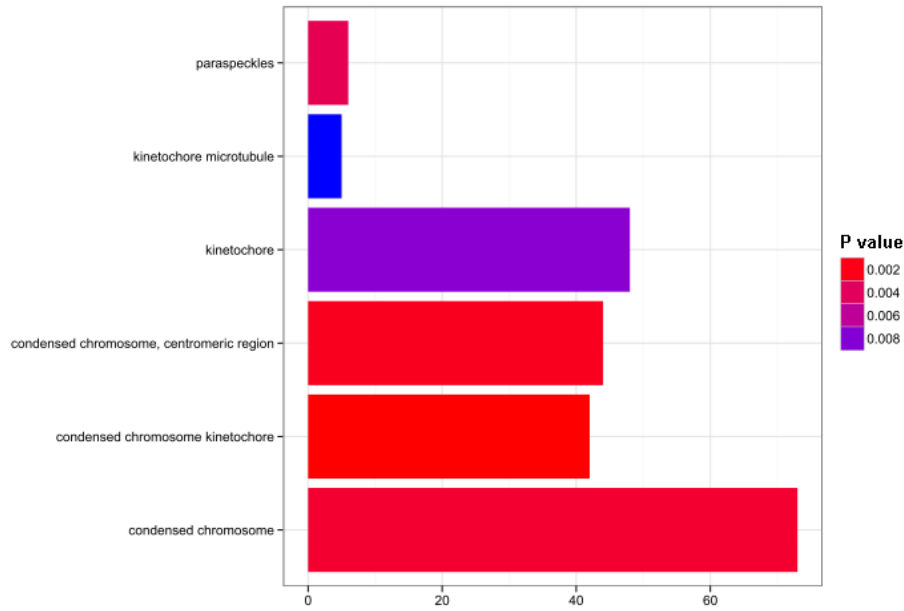
Figure 1. Classification of differentially expressed genes according to cellular component ontology. The colored entries are the significant aggregation (false discovery rate < 0.05) ones; P values are shown on the right side. The lower the P value, the more significant the enrichment is.
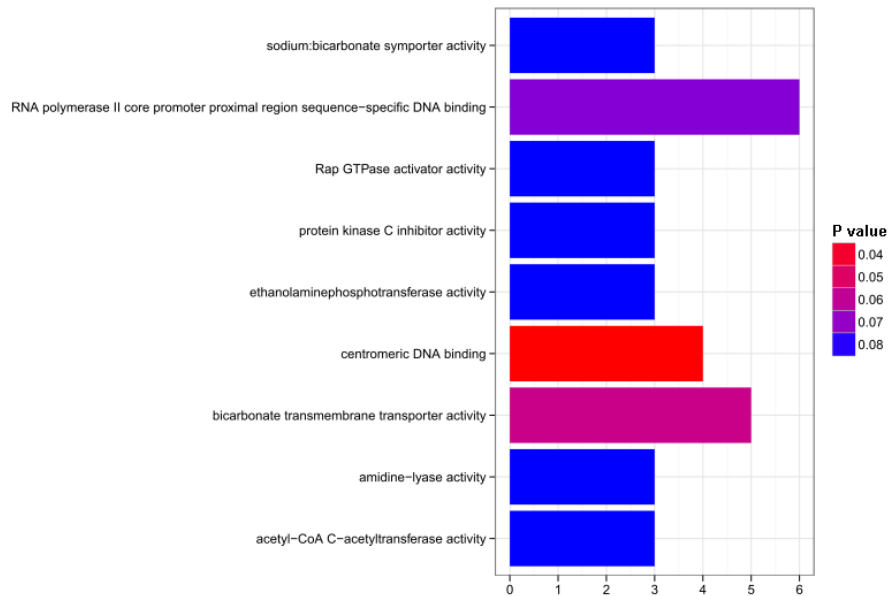

Figure 2. Classification of differentially expressed genes according to molecular function ontology. The colored entries are the significant aggregation (false discovery rate < 0.05) ones, the P values are shown on the right side. The lower the P value, the more significant the enrichment is.
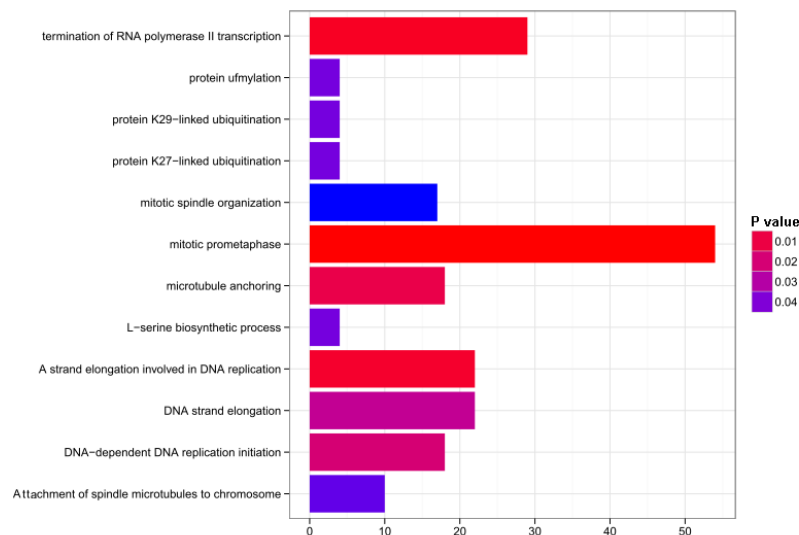
Figure 3. Classification of differentially expressed genes according to biological process ontology. The colored entries are the significant aggregation (false discovery rate < 0.05) ones; P values are shown on the right side. The lower the P value, the more significant the enrichment is.

Pathway enrichment analysis

For a deeper understanding of the different biological pathways in ADC and SCC, we chose KEGG Pathway cluster analysis to identify the significant pathways related to the DEGs. The top 16 pathways of ADC and SCC with a highly significant correlation are listed in Table 1. The DEGs were enriched in two more significant enrichment pathways: cell cycle (P = 1.38E-12) and DNA replication (P = 1.70E-05). The other significant pathways were spliceosome, p53 signaling pathway, oocyte meiosis and oocyte maturation pathways, ubiquitin-mediated proteolysis and lysosome, adherent junction, and tight junction.

Table 1. Enriched pathways of differentially expressed genes (DEGs) in adenocarcinoma and squamous cell carcinoma.

KEGG-ID Term	P value	DEGs (partially)
hsa04110 Cell cycle	1.38E-12	GSK3B, ATR, SKP2, CDK1, CDK2, SMC3, PLK1, CCND3
hsa03030 DNA replication	1.70E-05	RFC2, PRIM2, MCM4, MCM5
hsa03040 Spliceosome	9.42E-04	PRPF19, SRSF2, THOC4
hsa04115 p53 signaling pathway	0.001053	CDK1, CDK2, GTSE1, ATR
hsa04114 Oocyte meiosis	0.001301	IGF1R, SMC3, CCNB1, NA
hsa04914 Progesterone-mediated oocyte maturation	0.00294	IGF1R, CDK1, CDK2, BUB1, RPS6KA3, ADCY3, CCNB2
hsa04120 Ubiquitin-mediated proteolysis	0.003355	PRPF19, FANCL, UBE2A, PLAS2, SKP2, RNF7
hsa04520 Adherens junction	0.003981	IGF1R, TGFBR2, CTNND1
hsa04142 Lysosome	0.004596	IDS, ASAH1, SORT1, AP4S1
hsa00920 Sulfur metabolism	0.007886	SULT1A1, NA, PAPSS2
hsa00564 Glycerophospholipid metabolism	0.011143	PLD1, CDS1, PLA2G10, PCYT1B, DGKD, CHKA
hsa00740 Riboflavin metabolism	0.01468	RFK, ACP1, MTMR6
hsa00240 Pyrimidine metabolism	0.026111	NT5E, PNPT1, CAD
hsa04530 Tight junction	0.029106	CKD4, CASK, MPP5
hsa00900 Terpenoid backbone biosynthesis	0.033836	HMGCR, ACAT2, PDSS1, PDSS2, HMGCS1
hsa03440 Homologous recombination	0.034554	XRCC3, RAD51, RPA1, BRCA2, NA, BLM

DISCUSSION

The clinical therapy of ADC and SCC is usually based on histopathological features (Bishop et al., 2010). However, there is an urgent need to explore special treatment of these two sub-types. In this study, we defined a set of DEGs for ADC and SCC at the molecular level. The results showed that a total of 4124 DEGs were identified between ADC and SCC. With GO enrichment analysis, we discovered that these DEGs were associated with cell proliferation, cell signal transduction and metabolism. From the result of GO enrichment analysis in molecular function ontology, we identified differences between ADC and SCC, such as DNA replication. These findings are in accordance with previous studies. For example, functional enrichment analysis of genes both overexpressed and amplified in NSCLC has previously revealed a significant enrichment for DNA replication and repair (Dehan et al., 2007); ADC and SCC are characterized by distinct alterations in cell cycle regulation, DNA repair and metabolism (Daraselia et al., 2011, 2012).

To further identify the differentiating molecular mechanism of ADC and SCC, we identified 16 pathways. Table 1 shows the enrichment of the cyclin-dependent kinase (CDK) family genes in ADC and SCC. The genes were mainly involved in pathways associated with cell division and cell proliferation, such as cell cycle (CDK2, CDK4) and DNA replication, which played a major role in distinguishing ADC from SCC. In addition, several pathways involved in immune response (CDK1), cell signal transduction, and metabolism were also identified. These findings are consistent with previous studies. For instance, in NSCLC, H1 histone kinase activity related to CDK2 is also increased. Thus, CDK2 activity is negatively regulated by interactions with inhibitors, and it may play an important role in the proliferation of NSCLC (Kawana et al., 1998). With cDNA array and RT-PCR methods, Kettunen et al. (2004) screen out some upregulated and downregulated genes, which are overexpressed distinctly in ADC and SCC. Among them, CDK4 is only overexpressed in SCC and has been used to be a distinct chromosomal locus for SCC (Kettunen et al., 2004). We also observed the enrichment of SKP2 in ADC and SCC. As discussed elsewhere, SKP2 exhibits amplification status and is overexpressed in cell lines derived from NSCLC, so it may be involved in the progression of NSCLC. The increased expression of SKP2 is significantly correlated with the presence of SCC ($P = 0.037$). When the expression of SKP2 is low, the invasion and migration of NSCLC cells in culture are inhibited (Yokoi et al., 2004).

As expected, pathway enrichment analysis revealed that many metabolic pathways were dysregulated in ADC and SCC, such as p53 signaling pathway (Table 1). p53 signaling pathway controls cell-cycle progression via the regulation of genes, such as CDK-I p21 (WAF1/Cip1) and MDM2. In a molecular biology study of human papillomavirus infection, the p53 pathway was found to be inactivated, and there were several similarities and differences in SCC and ADC (Tjalma et al., 2005). On the basis of unique biological and pathological features, authors found that the most highly expressed genes in SCC and ADC encoded different kinds of proteins (Nacht et al., 2010). Although the clinical courses of ADC and SCC are similar, they can be distinguished through biological pathways (Nacht et al., 2001; Niklinska et al., 2001).

This study is important for the differentiation of ADC and SCC. As seen in Figures 1 to 3, our results indicated that proliferation capacity differed between ADC and SCC tumor cells. These results are in agreement with a previous study. Searching for a variety of key genes and signaling pathways in tumors has been the core of cancer research for over nearly

three decades. Thus, molecular signatures are needed to improve the diagnostic, prognostic, and therapeutic classification systems of NSCLC on the basis of molecular mechanisms (Agulló-Ortuño et al., 2010; Balko and Arteaga, 2012), especially in histologically heterogeneous tumors. Moreover, it is essential to use genomic signatures to optimize therapy; for example, EGFR amplification and increased gene copy number are more closely linked with SCC than ADC (Dacic et al., 2006; Sudo et al., 2013).

In conclusion, our study revealed that 4124 genes were differentially expressed in ADC and SCC. Pathways related to immune response, cell signal transduction, metabolism, cell division and cell proliferation were identified more in ADC than SCC. Therefore, these two diseases cannot be confused. At the same time, carrying out symptomatic treatment is particularly important. Our research may provide a potential strategy in the clinical therapy of ADC and SCC.

REFERENCES

American Thoracic Society (ATS) and European Respiratory Society (ERS) (1997). Pretreatment evaluation of non-small-cell lung cancer. *Am. J. Respir. Crit. Care Med.* 156: 320-332.

Agulló-Ortuño MT, Lopez-Rios F and Paz-Ares L (2010). Lung cancer genomic signatures. *J. Thorac. Oncol.* 5: 1673-1691.

Balko JM and Arteaga CL (2012). Molecular signatures of lung cancer: defining new diagnostic and therapeutic paradigms. *Mol. Diagn. Ther.* 16: 1-6.

Bishop JA, Benjamin H, Cholakh H, Chajut A, et al. (2010). Accurate classification of non-small cell lung carcinoma using a novel microRNA-based approach. *Clin. Cancer Res.* 16: 610-619.

Dacic S, Flanagan M, Cieply K, Ramalingam S, et al. (2006). Significance of EGFR protein expression and gene amplification in non-small cell lung carcinoma. *Am. J. Clin. Pathol.* 125: 860-865.

Daraselia N, Wang Y, Budoff A and Lituev A (2011). Pathway analysis of primary human non-small cell lung cancer (NSCLC). *J. Clin. Oncol.* (*Meeting Abstracts*) 29: 15-Suppl. 10573.

Daraselia N, Wang Y, Budoff A, Lituev A, et al. (2012). Molecular signature and pathway analysis of human primary squamous and adenocarcinoma lung cancers. *Am. J. Cancer Res.* 2: 93-103.

Dehan E, Ben-Dor A, Liao W, Lipson D, et al. (2007). Chromosomal aberrations and gene expression profiles in non-small cell lung cancer. *Lung Cancer* 56: 175-184.

Ferlay J, Autier P, Boniol M, Heanue M, et al. (2007). Estimates of the cancer incidence and mortality in Europe in 2006. *Ann. Oncol.* 18: 581-592.

Hou J, Aerts J, den Hamer B, van Ijcken W, et al. (2010). Gene expression-based classification of non-small cell lung carcinomas and survival prediction. *PLoS One* 5: e10312.

Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, et al. (2003). Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* 4: 249-264.

Kanehisa M and Goto S (2000). KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* 28: 27-30.

Kawana H, Tamaru J, Tanaka T, Hirai A, et al. (1998). Role of p27Kip1 and cyclin-dependent kinase 2 in the proliferation of non-small cell lung cancer. *Am. J. Pathol.* 153: 505-513.

Kettunen E, Anttila S, Seppanen JK, Karjalainen A, et al. (2004). Differentially expressed genes in nonsmall cell lung cancer: expression profiling of cancer-related genes in squamous cell lung cancer. *Cancer Genet. Cytogenet.* 149: 98-106.

Kim MJ, Shin HC, Shin KC and Ro JY (2013). Best immunohistochemical panel in distinguishing adenocarcinoma from squamous cell carcinoma of lung: tissue microarray assay in resected lung cancer specimens. *Ann. Diagn. Pathol.* 17: 85-90.

Kuner R, Muley T, Meister M, Ruschhaupt M, et al. (2009). Global gene expression analysis reveals specific patterns of cell junctions in non-small cell lung cancer subtypes. *Lung Cancer* 63: 32-38.

Nacht M, Dracheva T, Gao Y, Fujii T, et al. (2001). Molecular characteristics of non-small cell lung cancer. *Proc. Natl. Acad. Sci. U. S. A.* 98: 15203-15208.

Nacht M, Dracheva T, Sidransky D, Madden SL, et al. (2010). Molecular characteristics of non-small cell lung cancer. Google Patents.

Niklinska W, Chyczewski L, Laudanski J, Sawicki B, et al. (2001). Detection of P53 abnormalities in non-small cell lung cancer by yeast functional assay. *Folia Histochem. Cytobiol.* 39: 147-148.

R Development Core Team (2008). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna.

Scagliotti GV, Parikh P, von Pawel J, Biesma B, et al. (2008). Phase III study comparing cisplatin plus gemcitabine with cisplatin plus pemetrexed in chemotherapy-naive patients with advanced-stage non-small-cell lung cancer. *J. Clin. Oncol.* 26: 3543-3551.

Smyth GK (2004). Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat. Appl. Genet. Mol. Biol.* 3: Article3.

Sudo M, Chin TM, Mori S, Doan NB, et al. (2013). Inhibiting proliferation of gefitinib-resistant, non-small cell lung cancer. *Cancer Chemother. Pharmacol.* 71: 1325-1334.

Terry J, Leung S, Laskin J, Leslie KO, et al. (2010). Optimal immunohistochemical markers for distinguishing lung adenocarcinomas from squamous cell carcinomas in small tumor samples. *Am. J. Surg. Pathol.* 34: 1805-1811.

The Gene Ontology Consortium (2008). The Gene Ontology project in 2008. *Nucleic Acids Res.* 36: D440-D444.

Tjalma WA, Van Waes TR, Van den Eeden LE and Bogers JJ (2005). Role of human papillomavirus in the carcinogenesis of squamous cell carcinoma and adenocarcinoma of the cervix. *Best. Pract. Res. Clin. Obstet. Gynaecol.* 19: 469-483.

Yokoi S, Yasui K, Mori M, Iizasa T, et al. (2004). Amplification and overexpression of SKP2 are associated with metastasis of non-small-cell lung cancers to lymph nodes. *Am. J. Pathol.* 165: 175-180.

Yu G, Wang LG, Han Y and He QY (2012). clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS.* 16: 284-287.

Zhang YK, Zhu WY, He JY, Chen DD, et al. (2012). miRNAs expression profiling to distinguish lung squamous-cell carcinoma from adenocarcinoma subtypes. *J. Cancer Res. Clin. Oncol.* 138: 1641-1650.