MOLECULAR & CELLULAR TOXICOLOGY, Vol. 2, No. 4, 273-278, December 2006

Gene Expression Profile and Its Interpretation in Squamous Cell Lung Cancer

Dong Yoon Park[1], Jung Min Kim[1], Ja Eun Kim[1], Chang Hyuk Yoo[1], Han Yong Lee[1], Ji Young Song[1], Sang-Joon Hwang[1], Jae Cheal Yoo[1], Sung Han Kim[1], Jong Ho Park[2] & Jeong Ho Yoon[1]

[1]Digital Genomics, Inc., Room 805, Namsung Plaza (Ace IX), 345-30 Gasan-dong, Geumcheon-gu, Seoul 153-782, Korea
[2]Dept. of thoracic surgery, Korea Cancer Center Hospital, 215-4 Gongneung, Nowon-ku, Seoul 139-706, Korea
Correspondence and requests for materials should be addressed to J.H. Yoon (jhyoon@digital-genomics.co.kr)

Accepted 2 December 2006

Abstract

95 squamous cell lung carcinoma samples (normal tissue: 40 samples, tumor: 55 samples) were analyzed with 8 K cDNA microarray. 1-way ANOVA test was employed to select differentially expressed genes in tumor with FDR<0.01. Among the selected 1,655 genes, final 212 genes were chosen according to the expression fold change and used for following analysis. The expression of up-regulated 64 genes was verified with Reverse Transcription PCR and 10 genes were identified as candidates for SCC markers. In our opinion, those candidates can be exploited as diagnostic or therapeutic purposes. Gene Ontology (GO) based analysis was performed using those 212 genes, and following categories were revealed as significant biological processes: Immune response (GO: 0006955), antigen processing (GO: 0030333), inflammatory response (GO: 0006954), Cell adhesion (GO: 0007155), and Epidermis differentiation (GO: 0008544). Gene set enrichment analysis (GSEA) also carried out on overall gene expression profile with 522 functional gene sets. Glycolysis, cell cycle, K-ras and amino acid biosynthesis related gene sets were most distinguished. These results are consistent with the known characteristics of SCC and may be interconnected to rapid cell proliferation. However, the unexpected results from ERK activation in squamous cell carcinoma gripped our attention, and further studies are under progress.

Keywords: Microarray, Lung cancer, Expression profile

Lung cancer becomes a leading cause in worldwide cancer casualties. About one sixth of cancer death is caused by lung cancer. It is commonly classified into small cell lung cancer (SCLC) and non-small cell lung cancer (NSCLC), which is further categorized into adenocarcinoma (AC), squamous cell carcinoma (SCC), and large cell carcinoma (LC). These types of lung cancer show differences in the prognosis and treatment method as well as their histological characteristics. For NSCLC, surgery is thought to be the most effective treatment. However, many patients suffer from recurred disease and early diagnosis is the most successful way to increase the cure rate of lung cancer patients up to present.

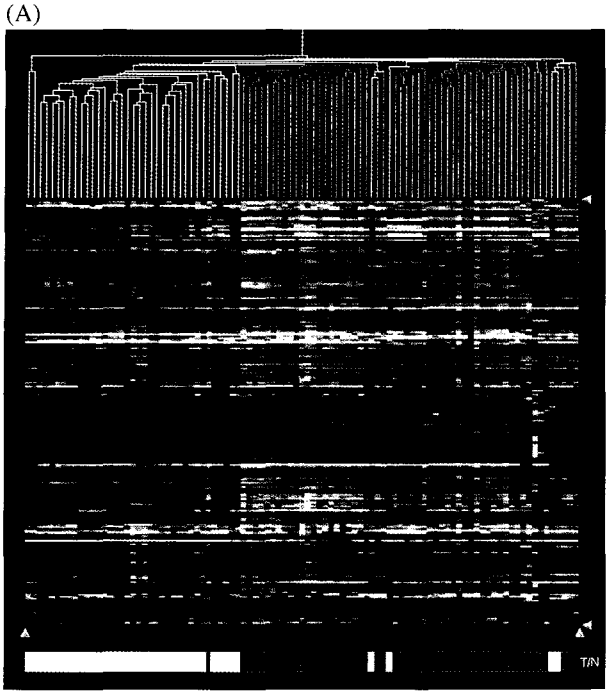
Several genes or proteins were identified as lung cancer markers by expression studies on limited number of targets[1-4]. Microarray analysis enabled screening of thousands of genes for lung cancer markers and revealed a number of genes with differential expression between lung cancer and normal tissues[5-7]. In principle, microarray analysis has potentials to provide not only the genes with differential expression but also the whole picture of gene expression that reflects physiological characteristics of the target cancer. However, most analysis reports generally focus on specific gene sets of interest because of our inability to interpret such huge information.

In this study, we present candidate genes for SCC specific marker through microarray analysis and Reverse Transcription PCR analysis. In addition, we provide the biological meanings associated with various gene expression signatures.

Gene Expression Profiles of Squamous Cell Lung Cancer and Normal Lung Tissue

Gene expression profiles of SCC and normal lung tissue samples were obtained through 8 K human cDNA microarray analysis. To obtain the overall patterns of gene expressions in SCC, we examined the gene expression data with hierarchical clustering and principal component analysis. Clustering analysis showed that samples are roughly divided into two large clusters of tumor and normal tissues (Fig. 1A). Tumor cluster may be subdivided into more minor-clusters. Other clinical factors such as stage, differentiation status, alcohol consumption, and smoking experience didn't show any significant association with the clustering pattern of tumor tissues. (χ^2 test, $P=$

(A)



(B)

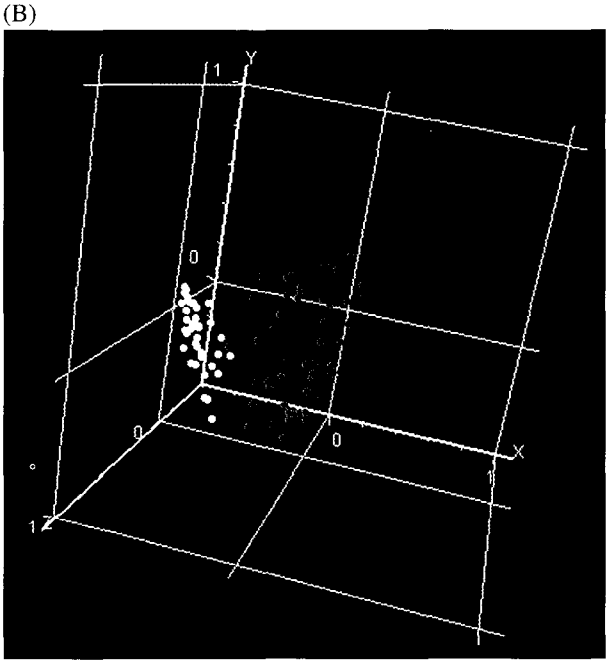


Fig. 1. Gene expression profile in SCC. (A) Hierarchical clustering analysis. 95 samples were divided into 2 clusters that represented normal tissue (yellow line in dendrogram) and SCC (red line in dendrogram), respectively. Gene expression levels of the differentially expressed 212 genes are depicted in the heat map under the dendrogram. (B) Principal component analysis in 3D scatter plot. As in hierarchical clustering result, 95 samples were clearly separated into 2 groups in 3D space. (yellow dot: normal tissue sample, red dot: SCC sample)

0.05, data not shown)

Gene expression differences between SCC and normal tissues were also evident in the principal component analysis that shows separation of SCC and normal samples in 3D space (Fig. 1B). These results indicate that our gene expression profile data may be useful to identify SCC marker genes and to understand the biological difference between SCC and normal lung tissues.

Genes with Differential Expression Between SCC and Normal Lung Tissue

We identified the genes with differential expression between SCC and normal lung tissue using 1-way ANOVA test with multiple testing correction (FDR $<$ 0.01). From the result of 1,655 genes, we only selected genes whose expressions were changed 2 fold above or 2 fold below in tumor compared to normal tissue. The final gene list contained 212 genes (64 up-regulated genes, 138 down-regulated genes). To understand cancer biology, both up-regulated and down-regulated genes are important. However, as a candidate for diagnostic marker or therapeutic target, up-regulated genes in cancer were preferred to down-regulated genes.

To verify the expression of the up-regulated genes in SCC, Reverse Transcription PCR was performed using eight pairs of SCC and normal tissue samples. Five sample pairs were obtained from stage I SCC patients, and the other three pairs were from stage III.

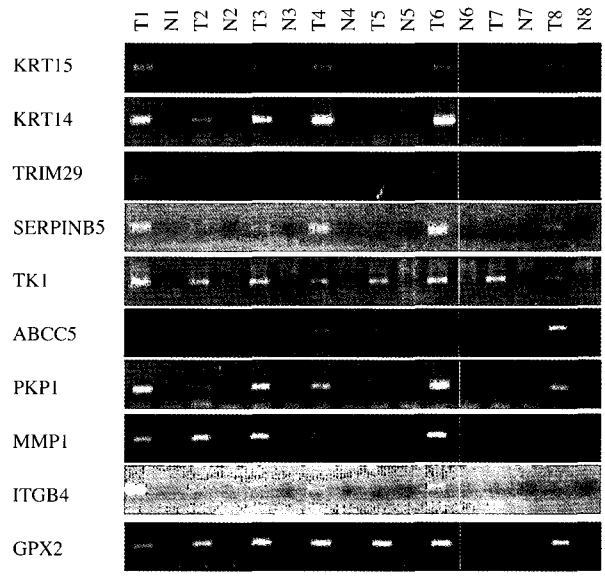


Fig. 2. Reverse transcription PCR results for SCC marker candidate genes. N1 means normal sample from patient No.1, and T1 means tumor sample from patient No.1.

The expression of the 10 genes was detected only in SCC samples but not in normal lung tissues (Fig. 2).

The tissue-specific expression patterns of 10 genes were checked out thoroughly by the public data search. KRT5 (keratin 15), KRT14 (keratin 14), TRIM29 (tripartite motif-containing 29), and SERPINB5 (serine (or cysteine) proteinase inhibitor, clade B, member 5) showed relatively high expression in prostate and thymus. TK1 (thymidien kinase, soluble), ITGB4 (integrin beta 4), ABCC5 (ATP-binding cassette, subfamily C, member 5), and GPX2 (glutathione peroxidase 2) showed none or low level of expression in normal lung tissue. The expressions of MMP1 (matrix metalloproteinase 1) and PKP1 (plakophilin 1) are consistently low in all examined tissues. Since their expression was strongly associated with squamous cell lung cancer, we propose these genes as candidates for SCC diagnostic markers.

Biological Processes Associated with Differentially Expressed Genes in SCC

In addition to the evaluating candidate marker genes, gene expression data may provide clues to understand the physiological characteristics of examined tissues. To systematically investigate the physiological characteristics of SCC at gene expression level, we looked into the biological roles of the selected 212 genes with differential expression between SCC and normal lung tissues. For this purpose, biological process information from Gene Ontology Consortium was used for functional categorization of genes and hypergeometric distribution was employed to calculate the significance of association. The biological processes that are associated with gene expression difference are summarized in Table 1.

The most interesting observation is the association of Immune response (GO: 0006955) category with the reduced gene expression in SCC. Among immune response related processes, antigen processing (GO: 0030333) and inflammatory response (GO: 0006954) shows strong relationship. Most of the down-regulated genes belonging to antigen processing category are encoding class II MHC molecules. Since class II MHC molecules are expressed only in antigen presenting cells, the number of antigen presenting cells in SCC tumor mass is smaller than that in normal lung tissue. And most genes in inflammatory response category are also known to be expressed in immune cells. In this category, genes encoding cell surface molecules involved in immune cell migration (FPRL1, ITGB2, and IL8RA), chemokine molecules (CCL18 and CXCL6), and well-known inflammation-related proteins (TNFRSF5, NFKB1, and IL1R1) were identified. Therefore, it is likely that number of immune cells and the resulting immune response is reduced in SCC and it may be beneficial for tumor cell survival.

Cell adhesion (GO: 0007155) is another biological process significantly associated with the reduced gene expression in SCC. This category also includes the genes involved in immune cell migration (AOC3, ITGB2, VCAM1, PECAM1, and ICAM1), which also suggests reduced immune response in SCC. Genes possibly related to cancer cell mobilization (CDH2, CDH5, NCAM1, ICAM1, and ICAM2) are also included in this category. In contrast, ITGB4 and ITGA6 that are known to promote carcinoma invasion[8] showed increased expression. Taken together, these gene expression profiles may represent increased motility of SCC cells and may contribute in metastasis.

Epidermis differentiation (GO: 0008544) is notable in categories associated with increased gene expression in SCC. It may represent squamous cell differentiation of SCC cells. Other categories also seem to be relevant to the characteristics of cancer cells such as uncontrolled proliferation, genome instability, and increased mutation. Therefore, biological processes

Table 1. The selected GO categories associated with differentially expressed genes.

GO ID	GO term	Ratio*	*P*-value	Ex.**
GO: 0008544	Epidermis development	5/8	0.000269	up
GO: 0006302	Double-strand break repair	2/11	0.005181	up
GO: 0000819	Sister chromatid segregation	2/12	0.006177	up
GO: 0030574	Collagen catabolism	2/13	0.007254	up
GO: 0007229	Integrin-mediated signaling pathway	2/25	0.025831	up
GO: 0030333	Antigen processing	6/18	0.000001	dn
GO: 0006952	Defense response	21/419	0.000001	dn
GO: 0006955	Immune response	20/389	0.000131	dn
GO: 0006954	Inflammatory response	8/105	0.001483	dn
GO: 0007155	Cell adhesion	12/295	0.019736	dn

*Ratio is the number of differentially expressed genes in a given category divided by the number of genes in that category on the whole microarray.
**In Ex. column, up means that category is associated with up-regulated genes and dn means its association with down-regulated genes.

Table 2. GSEA interpretation of gene expression profile in SCC. (FDR < 0.25).

Gene set	FDR
Glycolysis gluconeogenesis	0.031
Glutamine deprivation down-regulated genes	0.050
Mitochondrial genes	0.054
Purine metabolism	0.058
Proteasome degradation	0.064
Chemotaxis signalling	0.066
Pyrimidine metabolism	0.069
Electron transporter activity	0.074
Cell cycle	0.090
Leucine deprivation down-regulated genes	0.111
Oxidative Phosphorylation	0.118
K-ras signature genes	0.126
PGC related genes	0.140
mRNA splicing	0.170
Breast cancer estrogen signalling	0.228

associated with increased gene expression are consistent with known characteristics of SCC.

Gene Sets Associated with Overall Expression Profile of SCC

In addition to GO based analysis focused on differentially expressed genes in SCC, we also employed gene set enrichment analysis (GSEA) on whole expression data with default C_2 catalog of functional sets (522 gene sets) to identify meaningful gene sets. Both the methods depend on prior knowledge on genes and pathways, however GSEA is more flexible and comprehensive than conventional GO based analysis. GSEA results with FDR < 0.25 are summarized in Table 2.

As increased glycolysis is common characteristic of tumor, it's not surprising to find glycolysis related gene set at top rank and cell cycle related gene set is also expected. The up-regulation of sets of genes down-regulated by amino acid deprivation suggests the presence of increased amino acid biosynthesis. With gene set related to Ras activation, those results are likely related to the effects of rapid cell proliferation.

Discussion

We suggested 10 genes as candidates for SCC marker based on their expression and RT-PCR results. The overexpression of keratin genes (KRT14 and KRT15) and desmosomal proteins (PKP1) can be expected from the nature of SCC. Squamous differentiations manifested by the formation of keratin pearl and intercellular bridges are representative characteristics of SCC. Integrin $\alpha 6\beta 4$ promotes carcinoma invasion[8] and tumor angiogenesis[9]. Overexpression of MMP1 (matrix metalloproteinase 1), SERPINB5 (protease inhibitor 5), and TRIM29 (tripartite motif-containing 29) in NSCLC cancer was previously reported (Oncogene21_7749). Elevated serum and cytosolic thymidine kinase (TK1) activity have been observed in various human cancers including lung cancer[10]. ABCC5 (ATP-binding cassette, subfamily C, member 5; multidrug resistance associated protein 5) belongs to a multidrug resistance associated protein family that extrudes drugs out of the cells. Its expression is induced by chronic platinum drugs exposure in lung cancer[11].

GPX2 (glutathione peroxidase 2), and PKP1 (plakophilin 1) were newly identified in this study. GPX2 was previously identified as colon cancer marker in peripheral blood[12]. PKP1 is a desmosomal plaque protein involved in Intercellular bridging. However, overexpression of these genes in lung cancer has not been reported. Therefore, these genes are noble but reasonable candidates for SCC markers.

The remarkable gene expression characteristics of SCC learned by GO analysis were reduced immune response and cell adhesion. These attributes may be beneficial to cancer cell survival and facilitate cancer cell migration and metastasis. Though they seem to be general hallmarks of cancer rather than specific characteristics of squamous cell lung cancer.

In GSEA results, we found that K-ras related gene set is up-regulated in SCC, and it granted us interesting topic for further studies. Ras activation usually results in ERK activation through well-known Ras/MAPK signalling pathway. And it's reported that ERK is activated in about 35% Caucasian patients of NSCLC[13]. Because NSCLC of Caucasian is usually adenocarcinoma, we decided to confirm the ERK activation in squamous cell carcinoma of Korean patients. Unexpectedly we found that phosphorylated ERK was more evident in normal tissue than in SCC after small-scale pilot experiment. (data not shown) Now further studies are under progress to clarify the discrepancy of ERK activation.

Methods

Sample Preparation for cDNA Microarray Experiment

Lung cancer and normal lung tissue samples were obtained at the time of surgery. Consent was received for all patients who provided samples. Dissected mass was examined by a pathologist and divided into cancer and normal tissue. Tissues were cut into 1 cm^3 cubic and frozen in liquid nitrogen before 10 min

after removal from the patient.

Total RNA was extracted with TriZol reagent (InVitrogen) according to the manufacturers' recommendations. The quantity of prepared RNA was measured by using spectrophotometer. And integrity of RNA was checked with RNA 6000 Nano assay kit on the 2100 Bioanalyzer (Agilent).

cDNA Microarray Experiment

cDNA microarray used in this study (GenePlorer™ TwinChip™ human-8 K set 1, Digital Genomics, Inc., Seoul, Korea) is duplicated array where two identical 8K arrays are present at upper and lower region of a slide.

For parallel comparison of the gene expression profiles of 95 samples (normal: 40 samples, tumor: 55 samples), gene expression of each samples were compared with that of common reference sample. Common reference sample was acquired by pooling equal amount of RNA's from eight cell lines originated from lung (NCI-H23, NCI-H1299, NCI-H596, A-549, NCI-H358, NCI-H128, SK-LU-1, and Malme-3M).

And for hybridization, 20 μg of total RNA was reverse-transcribed with aminoallyl-modified dUTP and chemically coupled with fluorescent dyes. Samples from lung cancer patients were labeled with Cy5 and reference sample was labeled with Cy3. Pairs of labeled samples were combined and hybridized on the arrays. After washing, hybridized arrays were and scanned with ScanArray Lite (PerkinElmer) to produce Tiff image files. Tiff image files were quantified with GenePix 3.0 (Axon Instruments, Union City, CA).

Data Analysis

After quantification of the Tiff image data, the acquired GenePix result files were imported into GeneSpring 7.3 (Agilent). Data was normalized using intensity dependent (Lowess) normalization method[14], and the user-defined parameter f for the 'lowess' function was set to 20%. And only the genes present in at least 80% of samples were selected for further analysis.

Hierarchical clustering analysis was done on 95 samples with pearson correlation as similarity measure and average linkage as clustering algorithm. Principal component analysis (PCA) was also performed on samples using mean centering and scaling. For the class comparison between normal and tumor samples, 1-way ANOVA test was carried out to identify tumor specific genes. Both gene expression fold change (at least 2 fold above or below) and false discovery rate (FDR < 0.01) were considered for selecting differentially expressed genes.

The biological meaning of the selected genes was evaluated on the basis of annotation information including the Gene Ontology (GO) terms[15], Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways[16], and gene description from various public databases. To provide the probability measure for those biological processes, gene set enrichment analysis (GSEA)[17] and hypergeometric distribution based significance calculation[18] were applied.

Acknowledgements

This work was supported by the grant of the Ministry of Science and Technology, Republic of Korea (M10416130004-04N1613-00400).

References

1. Hibi, K. *et al.* PGP9.5 as candidate tumor marker for non-small cell lung cancer. *Am. J. Pathology* **155**, 711-715 (1999).
2. Brechot, J.M. *et al.* Diagnostic and prognostic value of Cyfra 21-1 compared with other tumor markers in patients with non-small cell lung cancer: a prospective study of 116 patients. *Eur. J. Cancer* **33**, 385-391 (1997).
3. Pastor, A. *et al.* Diagnostic value of SCC, CEA and CYFRA 21.1 in lung cancer: a Bayesian analysis. *Eur. Respir. J.* **10**, 603-609 (1997).
4. Morita, S. *et al.* Cancer cells overexpress mRNA of urokinase-type plasminogen activator, its receptor and inhibitors in human non-small cell lung cancer tissue: analysis by northern blotting and in situ hybridization. *Int. J. Cancer* **78**, 286-292 (1999).
5. Wang, T. *et al.* Identification of genes differentially over-expressed in lung squamous cell carcinoma using combination of cDNA subtraction and microarray analysis. Oncogene **19**, 1519-1528 (2000).
6. Petroziello, J. *et al.* Suppression subtractive hybridization and expression profiling identifies a unique set of genes overexpressed in non-small cell lung cancer. Oncogene **23**, 7734-7745 (2004).
7. Heighway, J. *et al.* Expression profiling of primary non-small cell lung cancer for target identification. Oncogene **21**, 7749-7763 (2002).
8. Shaw, L.M. *et al.* Activation of phosphoinositide 3-OH kinase by the α6β4 integrin promotes carcinoma invasion. Cell **91**, 949-960 (1997).
9. Nikolopoulos, S.N. *et al.* Integrin β4 signaling promotes tumor angiogenesis. Cancer Cell **6**, 471-483 (2004).
10. Romain, S. *et al.* DNA-synthesizing enzymes in breast cancer (thymidine kinase, thymidylate synthase

and thymidylate kinase): association with flow cytometric s-phase fraction and relative prognostic importance in node-negative premenopausal patients. *Int. J. Cancer* **95**, 56-61 (2001).
11. Oguri, T. *et al.* Increased expression of the MRP5 gene is associated with exposure to platinum drugs in lung cancer. *Int. J. Cancer* **86**, 95-100 (2000).
12. Solmi, R. *et al.* Search for epithelial-specific mRNAs in peripheral blood of patients with colon cancer by RT-PCR. *Int. J. Oncology* **25**, 1049-1056 (2004).
13. Vicent, S. *et al.* ERK1/2 is activated in non-small cell lung cancer and associated with advance tumours. *Br. J. Cancer* **90**, 1047-1052 (2004).
14. Yang, Y.H. *et al.* Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation. *Nucleic Acids Res.* **30**, e15 (2002).
15. Gene Ontology Consortium. The Gene Ontology (GO) database and information resource. *Nucleic Acids Res.* **32**, D258-D261 (2004).
16. Kanehisa, M. *et al.* From genomics to chemical genomics: new developments in KEGG. *Nucleic Acids Res.* **34**, D354-D357 (2006).
17. Subramanian, A. *et al.* Gene set enrichment analysis: A knowledge-based approach for interpreting genome -wide expression profiles. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 15545-15550 (2005).
18. Travazoie, S. *et al.* Systematic determination of genetic network architecture. *Nat. Genet.* **22**, 281-285 (1999).