

# ***In Silico* Comparative Genomic Analysis of Two Non-small Cell Lung Cancer Subtypes and their Potentials for Cancer Classification**

JINDONG LI<sup>1</sup>, DONGFANG LI<sup>1</sup>, XUDONG WEI<sup>1</sup> and YANHE SU<sup>2</sup>

<sup>1</sup>*Department of Thoracic Surgery, the Affiliated Cancer Hospital of Zhengzhou University & Henan Cancer Hospital, Zhengzhou, P.R. China*

<sup>2</sup>*Department of Thoracic Surgery, the Second Affiliated Hospital of Zhengzhou University, Zhengzhou, P.R. China*

**Abstract.** *Background/Aim:* Lung adenocarcinoma (AC) and squamous cell lung carcinoma (SCC) are two main subtypes of non-small cell lung cancer. In order to understand their biological differences, we conducted an *in silico* comparative genomic analysis of their expression profiles. *Materials and Methods:* We utilized the published microarray data of 18 SCC samples and 40 AC samples to discriminate genes differentially expressed in SCC and AC. Genes were employed to construct a functional module network and build a support vector machine classifier. Another set of published non-small cell lung cancer microarray data was used to test the predictive accuracy of support vector machine classifier. *Results:* Our analysis showed that SCC shows an elevated expression of genes related to cell division and DNA replication while AC presents an elevated expression of the genes related to protein transport and cell junction. ROC analysis demonstrates that the support vector machine classifier has a high classification accuracy for AC and SCC. *Conclusion:* AC and SCC are distinctively different in certain biological network modules. This proposes different pathological mechanisms involved in these two non-small cell lung cancer subtypes.

Lung cancer is one of major contributions of cancer-related death worldwide, bringing unbearable agony to patients and a hefty burden to health system. Since 2008, it has replaced

liver cancer as the most common malignant tumor in China (1). About 80% of lung cancers are non-small-cell lung cancers (NSCLC) (2), which are divided in three major subtypes: lung adenocarcinoma (AC), squamous cell lung carcinoma (SCC), and large cell lung carcinoma.

AC (nearly 40% of NSCLCs) and SCC (about 30% of NSCLCs) are more common than large cell lung carcinomas (2, 3). They compose over half of total lung cancer cases. AC is more associated to lung cancer patients without a history of smoking, although smoking is an important risk factor for lung cancer, especially in developed countries (4). SCC is more common in males than in females. It is closely associated with lung cancer patients having a smoking habit (5). These facts suggest that AC and SCC have different pathological mechanisms.

Microarray-based gene expression profiling has been used to describe the expression profiles of AC and SCC (6, 7). The genes differentially expressed between AC and SCC were identified and some of them are reported to be biomarkers for non-small cell lung cancer (8, 9). However, it remains unclear what the potential genes affecting the pathological process of AC or SCC are. So far, many studies only focused on a small portion of signature genes of AC or SCC and their possible roles in NSCLCs (10-13). Most information from AC and SCC microarray data is overlooked. The biological processes and molecular pathways hidden in these data are worth being uncovered.

The long-term survival rate of non-small cell lung cancer patients is abysmal (about 15% of 5-year overall survival and 8-10% of 10-year overall survival) (14, 15). The improvement of the patient's long-term survival rate requires for correct diagnosis of NSCLC subtype in early stage and proper treatment plan, both of which demand the understanding of the molecular basis underlying different NSCLC subtypes. Using the published AC and SCC microarray data (16), we conducted a comparative genomic analysis of AC and SCC expression profiles. We focused on

*Correspondence to:* Yanhe Su, Department of Thoracic Surgery, the Second affiliated hospital of Zhengzhou University, 2 Jingba Road, Zhengzhou, Henan Province, P.R. China, 450014. Tel: +86 037163620622. Fax: +86 037163620622, e-mail: yanhe\_su@163.com

**Key Words:** Lung adenocarcinoma, squamous cell lung carcinoma, microarray data, functional module network analysis, support vector machine classifier.

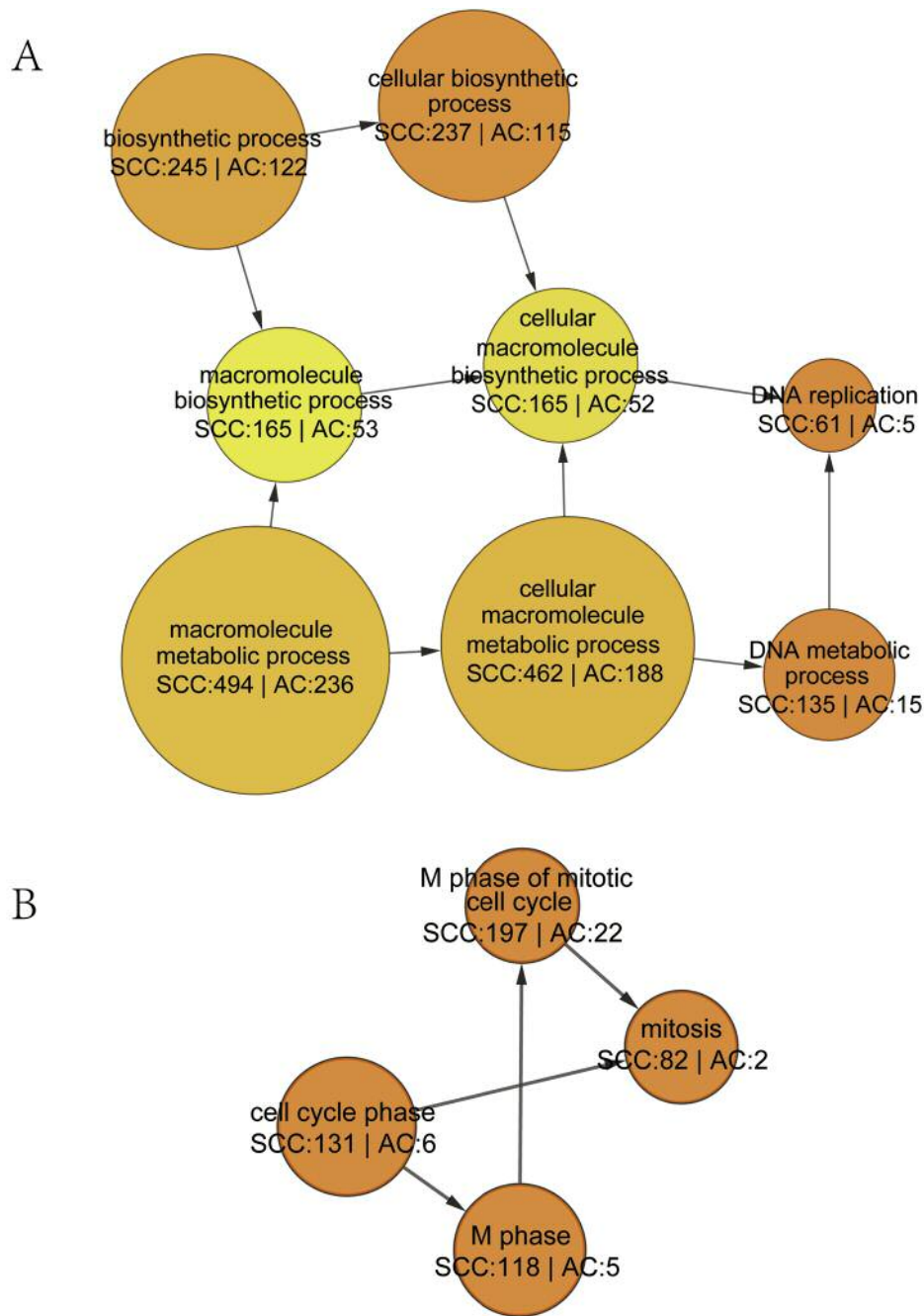


Figure 1. The highly interconnected modules extracted from the biological process network. The circles stand for module and the arrows stands for sub-network flow. The number beside SCC means the number of SCC up-regulated genes in the module and the number beside AC means the number of AC up-regulated genes in the module.

the differences between AC and SCC in their biological process, cellular component and molecular function networks. The major differences in their biological networks were identified in this study. In addition, we built a support vector machine (SVM) classifier based on genes

involved in these networks. It has a possible value in AC and SCC diagnosis. We also used another set of NSCLC microarray data to test for the accuracy of our classifier (8), which proved the high predictive accuracy of our SVM classifier.

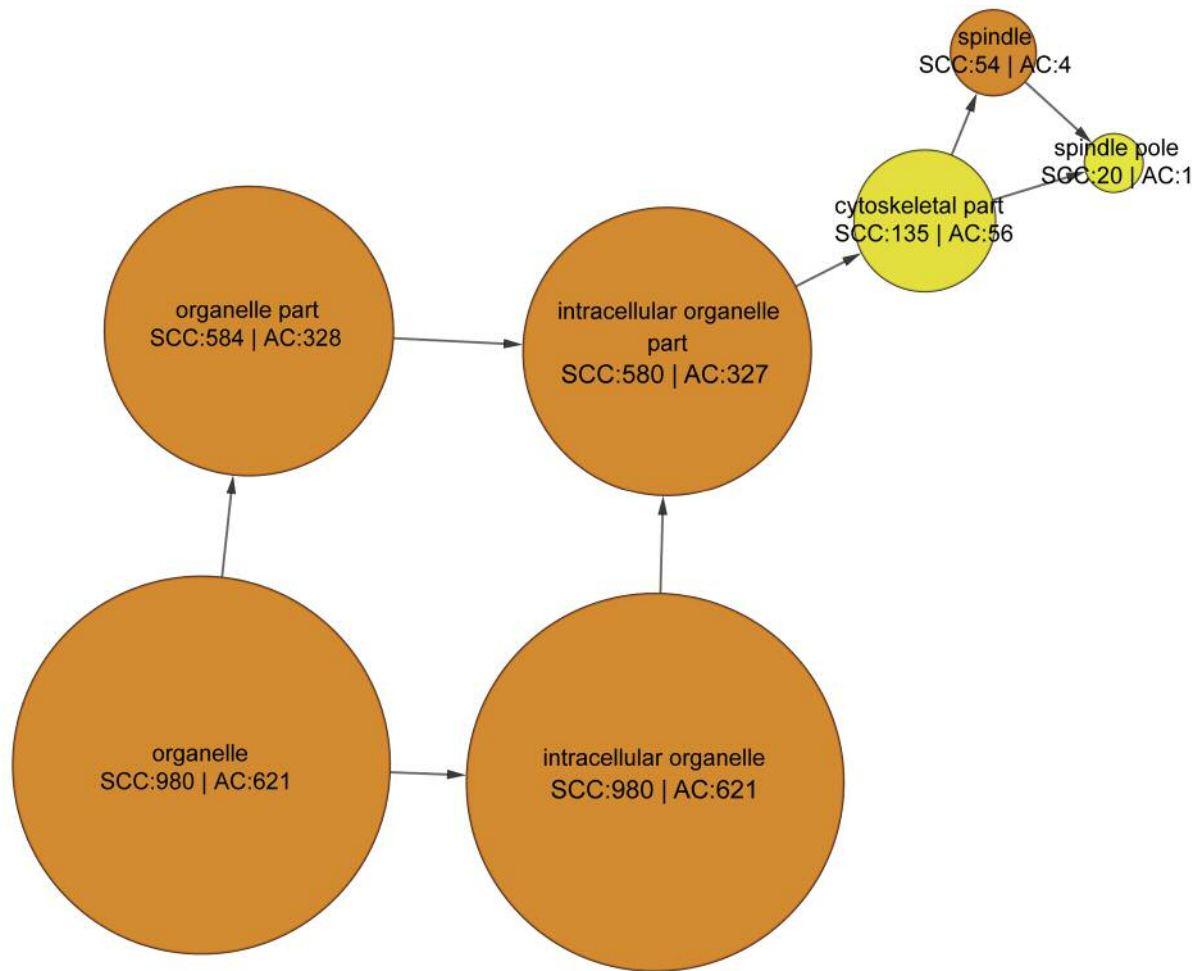


Figure 2. The highly interconnected modules extracted from the cellular component network. The circles stand for module and the arrows stand for sub-network flow. The number beside SCC indicates the number of SCC up-regulated genes in the module and the number beside AC means the number of AC up-regulated genes in the module.

## Materials and Methods

**Microarray data.** Microarray data including 18 SCC samples and 40 AC samples were downloaded from the GEO database (GSE10245). This data set was used for gene expression profile analysis and building of a support vector machine classifier. The test microarray data for support vector machine classifier were also downloaded from GEO database (GSE19804). It has 60 NSCLC samples (56 adenocarcinoma samples, 3 bronchioloalveolar carcinoma samples, and 1 squamous cell carcinoma sample). Both microarray data sets are normalized counts and were collected using Affymetrix Human Genome U133 Plus 2.0 Array (GPL570) (8, 16).

**Statistical methods for identifying differentially expressed genes.** One-way ANOVA test was employed to compare the mean normalized counts between AC and SCC samples in the first data set. The R package was used to perform statistical analysis and a  $p$ -value smaller than 0.01 was considered statistically significant.

**Functional enrichment analysis for differentially expressed genes.** We used the DAVID online resource to perform the gene-GO term enrichment analysis for the differentially expressed genes between AC and SCC (17, 18). The differentially expressed genes were classified as AC up-regulated genes and SCC up-regulated genes.

**Functional module network construction.** The differentially expressed genes mapped by DAVID were used for building a functional module network. BiNGO was used to cluster genes into functional modules and create the biological network based on these modules (19). MCODE was used to find the highly interconnected nodes in the networks (20).

**Building a support vector machine classifier.** The support vector machine classifier for AC and SCC was built with LIBSVM (21). The LIBSVM parameters were optimized with grid-search and 5-fold cross-validation. Only differentially expressed genes mapped by DAVID database were used as SVM features. We selected linear kernel for our SVM classifier, because the number of samples is

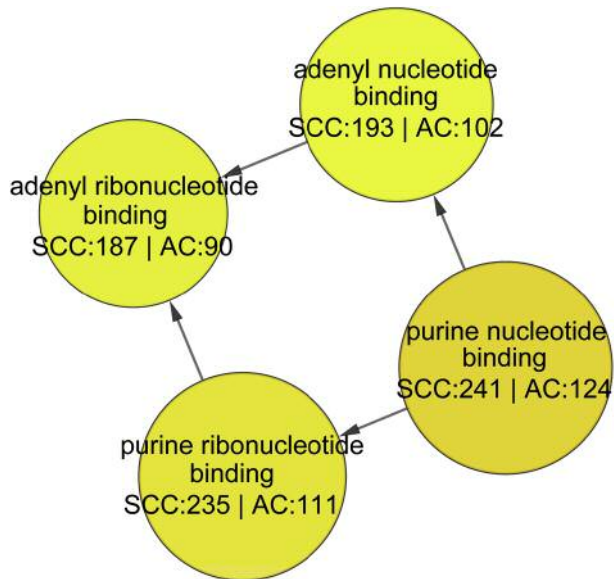


Figure 3. The highly interconnected modules extracted from the molecular function network. The circles stand for module and the arrows stands for sub-network flow. The number beside SCC means the number of SCC up-regulated genes in the module and the number beside AC means the number of AC up-regulated genes in the module.

greatly lesser than the number of features (22). Both the training and test data sets were normalized to the range [0, 1].

## Results

**Differentially expressed genes between AC and SCC.** Using the normalized gene expression data from the first microarray data set, we identified 3,544 differentially expressed genes between AC and SCC through ANOVA comparison ( $p$ -value  $< 0.01$ ). According to their expression levels in AC and SCC, we classified them into two categories: AC up-regulated genes and SCC up-regulated genes. Among 3,544 differentially expressed genes, 1,485 genes are up-regulated in AC while 2,059 genes are up-regulated in SCC.

**DAVID analysis of differentially expressed genes.** In order to understand the possible functions of these differentially expressed genes, we performed the GO term enrichment analysis for them using DAVID functional annotation software (17, 18). 2,857 out of 3,544 differentially expressed genes can be mapped in the DAVID database. 1,216 of them are AC up-regulated genes and 1,641 of them are SCC up-regulated genes. The functional enrichment result is shown in Table I. According to the result, the possible functions of differentially expressed genes between AC and SCC are mainly involved in cell cycle, condensed chromosome, purine nucleotide binding and DNA replication.

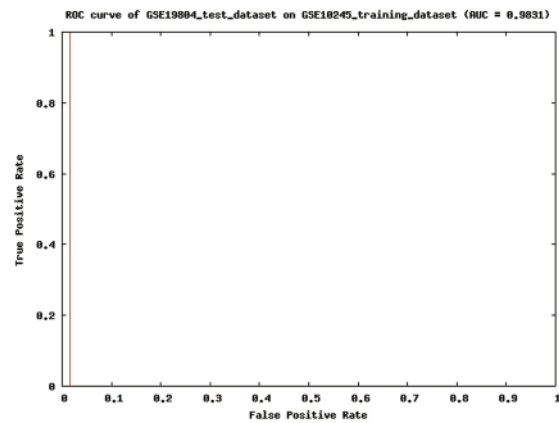


Figure 4. ROC analysis of the support vector machine classifier built on 2,857 mapped differentially expressed genes between AC and SCC.

**Functional module network analysis of differentially expressed genes.** Although the DAVID analysis of these differentially expressed genes yielded the information about their possible functions, it is still unclear how these functions contribute to the unique biological profile of AC or SCC. In order to investigate the relationship between these functions and the molecular basis of AC or SCC, we constructed biological interaction networks represented by these functions using a Cytoscape plugin, BiNGO (19). The networks of biological process, cellular component and molecular function were constructed. Nevertheless, these networks are too huge and too complex for us to analyze (data not shown). Thus, we used another Cytoscape plugin, MCODE, to refine these networks (20). The highly interconnected nodes were extracted from these networks with MCODE. These nodes are actually the important functional sub-networks in the whole network. We identified two functional sub-networks in biological process network, one sub-network in cellular component network and one sub-network in molecular function network (Figures 1, 2 and 3). The sub-networks in biological process network are responsible for DNA replication and mitosis (Figure 1A and B). They are mainly composed of SCC up-regulated genes. The highly interconnected modules in cellular component network are made of intracellular organelles which ultimately contribute to spindle apparatus, and the highly interconnected modules in molecular function network carry out the function of purine and adenyly nucleotide binding (Figure 2 and 3). About two-third of the genes involved in these modules are up-regulated in SCC. In all important sub-networks, there are more genes up-regulated in SCC than in AC. We also searched the modules with more AC up-regulated genes than SCC up-regulated ones and the result is shown in Table II. Although they are not highly

Table I. DAVID functional annotation analysis of differentially expressed genes between AC and SCC.

Category	Term	Benjamini-corrected FDR
Go: Biological process	M phase	1.0E-14
	Cell cycle	1.4E-14
	Cell cycle phase	4.7E-13
	M phase of mitotic cell cycle	9.1E-12
	Mitosis	2.4E-11
	Nuclear division	2.4E-11
	Cell cycle process	4.5E-11
	Organelle fission	6.7E-11
	Cell division	8.7E-11
Go: Cellular component	Condensed chromosome	1.0E-8
	Spindle	3.1E-8
	Nucleoplasm	4.9E-8
	Condensed chromosome kinetochore	3.7E-8
	Chromosome, centromeric region	7.2E-8
	Intracellular organelle lumen	7.3E-8
	Nuclear lumen	8.7E-8
Go: Molecular function	Purine nucleotide binding	2.1E-4
	Purine ribonucleotide binding	9.2E-4
	Ribonucleotide binding	1.1E-3
	Microtubule binding	9.9E-4
	Nucleotide binding	9.3E-4
	Tubulin binding	1.0E-3
	Adenyl nucleotide binding	1.0E-2
KEGG pathway	Cell cycle	1.5E-5
	DNA replication	1.9E-2
	Glycerophospholipid metabolism	1.3E-2

interconnected with other modules, these modules show that AC is more active in expression of the genes related to protein transport, endoplasmic reticulum, Golgi apparatus, and cell junction.

*The performance of support vector machine classifier.* The functional module network analysis shows that SCC up-regulated genes play a main part in DNA replication and cell cycle while AC up-regulated genes are mainly involved in protein transport and cell junction. They are useful features for separating these two major subtypes of NSCLC. We used 2,857 DAVID-mapped genes to build the support vector machine classifier for SCC and AC. 18 SCC samples and 40 AC samples from GSE10245 were used as training data set. We selected linear kernel function for building support vector machine classifier, because the number of genes is much larger than the number samples. First, we used the training data set to test the predictive actuary of our classifier. The result of training data set showed a classification accuracy of 100%. Secondly, we used another set of microarray data to test the predictive actuary of the classifier. This microarray data set includes 56 adenocarcinoma samples, 3 bronchioloaveolar carcinoma samples, and 1 squamous cell carcinoma sample. When tested with these samples, the

classifier showed an AUC score of 0.9831 in the ROC analysis (Figure 4). The classifier successfully identified AC and SCC samples, and predicted 2 bronchioloaveolar carcinoma samples as AC and 1 bronchioloaveolar carcinoma sample as SCC. Bronchioloaveolar carcinoma is usually considered as a subtype of lung adenocarcinoma (23, 24), so the predictive accuracy of this classifier on AC is 98.3% (58/59). This result proves its possible value in AC and SCC diagnosis.

## Discussion

The identification of differentially expressed genes between AC and SCC could help elucidate their different oncogenic mechanisms. Our study shows that there are more SCC up-regulated genes than AC up-regulated ones in the differentially expressed genes. It suggests that the pathological process underlying SCC is more complex than that of AC. The pathogenesis of SCC might need more steps of somatic mutation which, in turn, recruited more abnormally-expressed genes. The clinical statistics show that SCC is less common than AC among NSCLCs (2, 3), which circumstantially support the statement above. The males with a history of tobacco use are more susceptible to SCC while

Table II. The functional modules in biological networks with more AC up-regulated genes than SCC up-regulated ones.

Network category	Module	The number of AC up-regulated genes vs. the number of SCC up-regulated ones
Biological process	Establishment of protein localization	84 vs. 73
	Protein transport	84 vs. 71
Cellular component	Endomembrane system	168 vs. 124
	Organelle membrane	194 vs. 160
	Endoplasmic reticulum	123 vs. 86
	Golgi membrane	61 vs. 43
	Endoplasmic reticulum part	87 vs. 57
	Endoplasmic reticulum membrane	80 vs. 51
	Nuclear membrane-endoplasmic reticulum network	82 vs. 51
	Golgi apparatus part	68 vs. 47
	Golgi apparatus	109 vs. 68
	Subsynaptic reticulum	88 vs. 59
	Apicolateral plasma membrane	18 vs. 10
	Apical junction complex	18 vs. 9
	Anchoring junction	21 vs. 18
	Occluding junction	15 vs. 8
	Tight junction	15 vs. 8

AC is the most common type of lung cancer in non-smokers (4, 5). Carcinogens like benzopyrene in tobacco smoke could cause and facilitate the mutation of normal cells into SCC cancer cells. AC is more common among lung cancer patients without smoking habit. Although the significance of genetic factors in AC development is unknown, they definitely play a part in the carcinogenesis of AC (25, 26).

In the present study, the majority of the identified differentially expressed genes are annotated by DAVID database. The DAVID analysis shows that their biological process, cellular component, molecular function, and KEGG pathway are mainly involved in cell cycle, condensed chromosome, purine nucleotide binding, and DNA replication, respectively. Since one major feature of cancer is uncontrolled proliferation, it is expected that their functions are somewhat related to cell division. The functional module network analysis of these genes revealed more detailed information about the difference between SCC and AC. In all highly interconnected modules in biological process, cellular component, and molecular function network, there are more SCC up-regulated genes than AC up-regulated ones. Their elevated expression in SCC indicates that SCC has a faster rate of DNA replication and cell division than AC, which is consistent with the former study (27). In cellular component sub-network, the organelle genes form a sub-network which finally contributes to spindle apparatus. As an indispensable cellular structure for cell division, the high expression level of spindle genes in SCC confirmed its hyperactivity in cell division. The modules in molecular function sub-network are responsible for nucleotide and ribonucleotide binding. The fact that SCC

has more up-regulated genes in these modules suggests its fast rate of DNA synthesis. On the other hand, AC up-regulated genes are more concentrated in the modules related to protein transport, endoplasmic reticulum, Golgi apparatus, and cell junction. While these modules do not form an interconnected sub-network, they still suggest that AC might be a result of abnormal expression of cell-cell interaction genes and cell junction genes. The functional module network analysis indicates the different molecular basis for AC or SCC carcinogenesis. Although the crucial genes influencing the process of AC or SCC carcinogenesis remain unknown, our study proposes that different drugs and different treatment strategies should be considered for different NSCLC subtypes in lung cancer therapy.

The SVM classifier built on the DAVID-mapped genes shows a high accuracy for identifying AC samples. Due to the limited number of SCC samples in test data set, its predictive accuracy for SCC still needs to be evaluated. We have confidence in the predictive power of our classifier. At least, when tested with the training data set, its predictive accuracy for both AC and SCC is 100%. Its performance on bronchioloalveolar carcinoma is intriguing. It classified 2 bronchioloalveolar carcinoma samples as AC and 1 bronchioloalveolar carcinoma sample as SCC. Bronchioloalveolar carcinoma is commonly regarded as a subtype of AC (23, 24). However, our result suggests that it might be not the case. If our result were correct, bronchioloalveolar carcinoma should be viewed as a mix of several lung cancer variants rather than a single variant of lung cancer. Further studies are required to elucidate the pathological classification of this less common type of lung cancer.

In conclusion, our analysis showed that 3,544 genes are differentially expressed between AC and SCC. SCC has greater number of up-regulated, differentially expressed, genes than AC. The functional enrichment analysis shows that these genes are mainly involved in cell cycle and DNA replication, and the functional module network analysis indicates that SCC and AC have different molecular bases and biological profiles. SCC has an elevated expression of the genes related to cell division and DNA replication while AC has an elevated expression of the genes related to protein transport and cell junction. These results indicate the different pathological mechanisms of SCC and AC. Further investigations are required to identify the genes and the molecular pathways controlling these mechanisms, if we want to exploit them as bases for NSCLC treatment. We also used the differentially expressed genes to build a support vector machine classifier for SCC and AC. It demonstrates the high predictive accuracy for AC and has a potential value in NSCLC diagnosis.

## References

- 1 She J, Yang P, Hong Q and Bai C: Lung cancer in China: challenges and interventions. *Chest* 143: 1117-1126, 2013.
- 2 Nugent WC, Edney MT, Hamneress PG, Dain BJ, Maurer LH and Rigas JR: Non-small cell lung cancer at the extremes of age: impact on diagnosis and treatment. *The Annals of thoracic surgery* 63: 193-197, 1997.
- 3 Kenfield SA, Wei EK, Stampfer MJ, Rosner BA and Colditz GA: Comparison of aspects of smoking among the four histological types of lung cancer. *Tobacco control* 17: 198-204, 2008.
- 4 Subramanian J and Govindan R: Lung cancer in never smokers: a review. *Journal of clinical oncology: official journal of the American Society of Clinical Oncology* 25: 561-570, 2007.
- 5 Wood ME, Kelly K, Mullineaux LG and Bunn PA Jr.: The inherited nature of lung cancer: a pilot study. *Lung Cancer* 30: 135-144, 2000.
- 6 Kikuchi T, Daigo Y, Katagiri T, Tsunoda T, Okada K, Kakiuchi S, Zembutsu H, Furukawa Y, Kawamura M, Kobayashi K, Imai K and Nakamura Y: Expression profiles of non-small cell lung cancers on cDNA microarrays: identification of genes for prediction of lymph-node metastasis and sensitivity to anti-cancer drugs. *Oncogene* 22: 2192-2205, 2003.
- 7 Tanney A, Oliver GR, Farztdinov V, Kennedy RD, Mulligan JM, Fulton CE, Farragher SM, Field JK, Johnston PG, Harkin DP, Proutski V and Mulligan KA: Generation of a non-small cell lung cancer transcriptome microarray. *BMC medical genomics* 1: 20, 2008.
- 8 Lu TP, Tsai MH, Lee JM, Hsu CP, Chen PC, Lin CW, Shih JY, Yang PC, Hsiao CK, Lai LC, and Chuang EY: Identification of a novel biomarker, SEMA5A, for non-small cell lung carcinoma in nonsmoking women. *Cancer epidemiology, biomarkers & prevention: a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology* 19: 2590-2597, 2010.
- 9 Botling J, Edlund K, Lohr M, Hellwig B, Holmberg L, Lambe M, Berglund A, Ekman S, Bergqvist M, Ponten F, Konig A, Fernandes O, Karlsson M, Helenius G, Karlsson C, Rahnenfuhrer J, Hengstler JG and Micke P: Biomarker discovery in non-small cell lung cancer: integrating gene expression profiling, meta-analysis, and tissue microarray validation. *Clinical cancer research: an official journal of the American Association for Cancer Research* 19: 194-204, 2013.
- 10 Au NH, Gown AM, Cheang M, Huntsman D, Yorlida E, Elliott WM, Flint J, English J, Gilks CB and Grimes HL: P63 expression in lung carcinoma: a tissue microarray study of 408 cases. *Applied immunohistochemistry & molecular morphology: AIMM/official publication of the Society for Applied Immunohistochemistry* 12: 240-247, 2004.
- 11 Shi I, Hashemi Sadraei N, Duan ZH and Shi T: Aberrant signaling pathways in squamous cell lung carcinoma. *Cancer informatics* 10: 273-285, 2011.
- 12 Ibrahim R, Matsubara D, Osman W, Morikawa T, Goto A, Morita S, Ishikawa S, Aburatani H, Takai D, Nakajima J, Fukayama M, Niki T and Murakami Y: Expression of PRMT5 in lung adenocarcinoma and its significance in epithelial-mesenchymal transition. *Human pathology* 45: 1397-1405, 2014.
- 13 Seike M, Yanaihara N, Bowman ED, Zanetti KA, Budhu A, Kumamoto K, Mechanic LE, Matsumoto S, Yokota J, Shibata T, Sugimura H, Gemma A, Kudoh S, Wang XW and Harris CC: Use of a cytokine gene expression signature in lung adenocarcinoma and the surrounding tissue as a prognostic classifier. *Journal of the National Cancer Institute* 99: 1257-1269, 2007.
- 14 Mountain CF, Lukeman JM, Hammar SP, Chamberlain DW, Coulson WF, Page DL, Victor TA and Weiland LH: Lung cancer classification: the relationship of disease extent and cell type to survival in a clinical trials population. *Journal of surgical oncology* 35: 147-156, 1987.
- 15 Naruke T, Goya T, Tsuchiya R and Suemasu K: Prognosis and survival in resected lung carcinoma based on the new international staging system. *The Journal of thoracic and cardiovascular surgery* 96: 440-447, 1988.
- 16 Kuner R, Muley T, Meister M, Ruschhaupt M, Bunes A, Xu EC, Schnabel P, Warth A, Poustka A, Sultmann H and Hoffmann H: Global gene expression analysis reveals specific patterns of cell junctions in non-small cell lung cancer subtypes. *Lung Cancer* 63: 32-38, 2009.
- 17 Huang da W, Sherman BT and Lempicki RA: Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic acids research* 37: 1-13, 2009.
- 18 Huang da W, Sherman BT and Lempicki RA: Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature protocols* 4: 44-57, 2009.
- 19 Maere S, Heymans K and Kuiper M: BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. *Bioinformatics* 21: 3448-3449, 2005.
- 20 Rivera CG, Vakil R and Bader JS: NeMo: Network Module identification in Cytoscape. *BMC bioinformatics* 11(Suppl 1): S61, 2010.
- 21 Chih-Chung Chang C-JL: LIBSVM: a library for support vector machines. *ACM Transactions on Intelligent Systems and Technology* 2: 1-27, 2011.
- 22 Rong-En Fan K-WC, Cho-Jui Hsieh, Xiang-Rui Wang, Chih-Jen Lin: LIBLINEAR: A library for large linear classification. *Journal of Machine Learning Research*: 1871-1874, 2008.

- 23 Dermer GB: Origin of bronchioloalveolar carcinoma and peripheral bronchial adenocarcinoma. *Cancer* 49: 881-887, 1982.
- 24 Raz DJ, Kim JY and Jablons DM: Diagnosis and treatment of bronchioloalveolar carcinoma. *Current opinion in pulmonary medicine* 13: 290-296, 2007.
- 25 Hsiung CA, Lan Q, Hong YC, Chen CJ, Hosgood HD, Chang IS, Chatterjee N, Brennan P, Wu C, Zheng W, Chang GC, Wu T, Park JY, Hsiao CF, Kim YH, Shen H, Seow A, Yeager M, Tsai YH, Kim YT, Chow WH, Guo H, Wang WC, Sung SW, Hu Z, Chen KY, Kim JH, Chen Y, Huang L, Lee KM, Lo YL, Gao YT, Liu L, Huang MS, Jung TH, Jin G, Caporaso N, Yu D, Kim CH, Su WC, Shu XO, Xu P, Kim IS, Chen YM, Ma H, Shen M, Cha SI, Tan W, Chang CH, Sung JS, Zhang M, Yang TY, Park KH, Yuenger J, Wang CL, Ryu JS, Xiang Y, Deng Q, Hutchinson A, Kim JS, Cai Q, Landi MT, Yu CJ, Tucker M, Hung JY, Lin CC, Perng RP, Boffetta P, Chen CY, Chen KC, Yang SY, Hu CY, Chang CK, Fraumeni JF Jr., Chanock S, Yang PC, Rothman N and Lin D: The 5p15.33 locus is associated with risk of lung adenocarcinoma in never-smoking females in Asia. *PLoS genetics* 6:2010.
- 26 Hosgood HD, 3rd, Wang WC, Hong YC, Wang JC, Chen K, Chang IS, Chen CJ, Lu D, Yin Z, Wu C, Zheng W, Qian B, Park JY, Kim YH, Chatterjee N, Chen Y, Chang GC, Hsiao CF, Yeager M, Tsai YH, Wei H, Kim YT, Wu W, Zhao Z, Chow WH, Zhu X, Lo YL, Sung SW, Chen KY, Yuenger J, Kim JH, Huang L, Chen YH, Gao YT, Huang MS, Jung TH, Caporaso N, Zhao X, Huan Z, Yu D, Kim CH, Su WC, Shu XO, Kim IS, Bassig B, Chen YM, Cha SI, Tan W, Chen H, Yang TY, Sung JS, Wang CL, Li X, Park KH, Yu CJ, Ryu JS, Xiang Y, Hutchinson A, Kim JS, Cai Q, Landi MT, Lee KM, Hung JY, Tucker M, Lin CC, Ren Y, Perng RP, Chen CY, Jin L, Chen KC, Li YJ, Chiu YF, Tsai FY, Yang PC, Fraumeni JF, Jr., Seow A, Lin D, Zhou B, Chanock S, Hsiung CA, Rothman N and Lan Q: Genetic variant in TP63 on locus 3q28 is associated with risk of lung adenocarcinoma among never-smoking females in Asia. *Human genetics* 131: 1197-1203, 2012.
- 27 Inamura K, Fujiwara T, Hoshida Y, Isagawa T, Jones MH, Virtanen C, Shimane M, Satoh Y, Okumura S, Nakagawa K, Tsuchiya E, Ishikawa S, Aburatani H, Nomura H and Ishikawa Y: Two subclasses of lung squamous cell carcinoma with different gene expression profiles and prognosis identified by hierarchical clustering and non-negative matrix factorization. *Oncogene* 24: 7105-7113, 2005.

*Received August 15, 2014*

*Revised September 29, 2014*

*Accepted October 1, 2014*