European Review for Medical and Pharmacological Sciences 2012; 16: 1582-1587

Bioinformatics analysis of two microarray gene-expression data sets to select lung adenocarcinoma marker genes

X. WU[1], W. ZANG[2], S. CUI[1], M. WANG[1]

[1]Department of Cardiothoracic Surgery, Shanghai 10th People's Hospital, Shanghai, China
[2]Department of Biology, Shanghai Jiaotong University, Shanghai, China
Xintian Wu and Weidong Zang worked equally to this paper

Abstract. – BACKGROUND: Lung adenocarcinoma (LAC) is the most frequent histologic type of lung cancer and rates of adenocarcinoma are increasing in most countries. Recently, several molecular markers have been identified to predict LAC. However, more prognostic makers and the underlying role of those makers are still imperative.

AIM: In this study, our objective was to identify a set of discriminating genes that can be used for characterization and prediction of response to LAC.

MATERIALS AND METHODS: Using the bioinformatics analysis method, we merged two LAC datasets-GSE2514 and GSE7670 to find novel target genes and pathways to explain the pathogenicity.

RESULTS: The results showed that EDNRB (endothelin receptor type B), ADRB2 (beta-adrenergic receptor), S1PR1 (sphingosine-1-phosphate receptor 1), P2RY14 (PsY purinoceptor 14), LEPR (leptin-receptor), GHR (growth hormone receptor), PPM1D (protein phosphatase-1D), and GADD45B (growth arrest and DNA-damage-inducible, beta) have high degrees in response to LAC. Additionally, EDNRB, ADRB2, S1PR1, P2RY14, LEPR, and GHR may be involved in LAC through Neuroactive ligand-receptor interaction, but PPM1D and GADD45B may be through p53 signaling pathway. Some of our prediction had been demonstrated by previous reports, such as ADRB2, S1PR1, GHR, PPM1D, and GADD45B. Therefore, we hope our study could lay a basis for further study of other target genes, such as EDNRB, P2RY14, and LEPR.

CONCLUSIONS: It is effective to identify potential molecular marker for LAC and predict their underlying functions by bioinformatics analysis and graph clustering method. However, further experiments are still indispensable to confirm our conclusion.

Key Words:
Graph-cluster, Lung adenocarcinoma, Bioinformatics analysis.

Introduction

Lung cancer is the most common cancer as well as the leading cause of cancer-related deaths in most countries. Among the four major histological types, adenocarcinoma is the predominant type in recent years[1]. Despite advances in treatment, the 5 year overall survival rate is approximately 15.7%[2]. Therefore, it is essential to identify specific molecular markers for lung adenocarcinoma (LAC) to improve early prediction[3].

To date, several tumor markers have been studied for detection of LAC, such as Ets-1[4], Kruppel-like factor 6 (KLF6)[5], Eukaryotic initiation factor 4E (eIF4E)[6], Nectin-like molecule-5 (Necl-5)[7], and Histone deacetylase inhibitors (HDACis)[8,9], which all have been demonstrated as poor prognostic makers for LAC. The higher levels of their expression were correlated with poorer histopathological differentiation, higher pathological stage and clinical stage, and a higher incidence of hematogenous metastasis. And the 5 year disease-free survival rate in patients with positive their overexpression was significantly poorer than among those expressing lower levels. In addition, there are some molecular markers identified for good prognosis of LAC, such as Runt-related transcription factor 3 (RUNX3)[10], estrogen receptor[11], and some chemokine receptors[12,13]. The high expression of them was significantly correlated with an increasing disease-free survival in LAC patients.

However, more molecular markers of lung adenocarcinoma are still on demand. Bioinformatics analysis provides a powerful tool for analyzing microarray experiments by combining data from multiple studies. The Bioconductor package RankProd provides a new and intuitive tool for this purpose in detecting differentially expressed genes under two experimental conditions[14]. Thus, in this study, we aimed to use the meta-analysis to detect more differentially expressed genes and then used the graph-clustering approach to further identify gene expression profiles that distinguish lung tumors from normal

Corresponding Author: Shitao Cui, MD; e-mail: cuishitaocst@gmail.com

lung samples. And the relevant pathways in the cluster are also analyzed to explain potential mechanisms in response to LAC.

Materials and Methods

Bioinformatics Analysis for the Expression Profiles and Differentially Expressed Genes (Degs) Analysis

Two LAC related expression profiles GSE2514 and GSE7670 were obtained from a public functional genomics data repository GEO (http://www.ncbi.nlm.nih.gov/geo/) which are based on the Affymetrix Human Genome U95 Version 2 Array and Affymetrix Human Genome U133A Array, separately. In this study, total 48 lung tumors and 47 controls were selected respectively to identify different expressed genes.

Statistical Analysis

For the GSE2514 and GSE7670 datasets, the RankProd package was used to identify DEGs. The DEGs only with a percentage of false-positives (PFP)[14] ≤ 0.05 were considered differentially expressed between treatments and controls.

To demonstrate the potential connection between DEGs, the Spearman rank correlation (r) was used for comparative target genes correlations. The significance level was set at $r > 0.9$ and fdr < 0.05. All statistical tests were performed with the R program (http://www.r-project.org/).

Network Analyses and Graph Clustering

To identify co-expressed groups we used DPClus[15]. A graph clustering algorithm that can extract densely connected nodes as a cluster. It is based on density-and periphery tracking of clusters. DPClus is freely available from http://kanaya.naist.jp/DPClus/. In this study, we used the overlapping-mode with the DPClus settings. We set the parameter settings of cluster property cp; density values were set to 0.5[16].

Pathway Enrichment Analysis

The Pathway[17] database records networks of molecular interactions in the cells, and variants of them specific to particular organisms (http://www.genome.jp/kegg/). The DAVID[18] was used to identify over-represented pathways. The *p*-value < 0.01 was as the significant.

Results

Differently Genes Selection and A Correlation Network Construction

In order to screen the differentially expressed genes, we downloaded publicly available microarray data sets GSE2514 and GSE7670 from GEO. In the microarray analysis, the differentially expressed genes (DEGs) with the fold change > 2 and *p*-value < 0.05 were selected. 340 genes from GSE2514 and 1585 genes from GSE7670 were selected as DEGs. Using the RankProd packages for meta-analysis, 5 up-regulated genes and 1025 down-regulated genes with a percentage of false-positives (PFP) < 0.01 were considered differentially expressed. At last, 1030 DEGs were collected after the meta-analysis.

To get the relationships among DEGs, the co-expressed value r = 0.9 and corrected *p*-value = 0.01 was chosen as the cut off value. Finally, 4509 relationships among 377 DEGs were constructed a correlation network (Figure 1).

Graph Clustering Identifies Modules Significantly Enriched in Biochemical Pathways

At $r > 0.9$, DPClus[15] identified 4 clusters, in the correlation network (Figure 1) for response to LAC; They ranged in size from 17 to 96 genes. Each cluster has the connection between each other. Graph clustering results are presented in Figure 2. To assess the significance of the clusters, we used the over-represented pathways (also called pathway enrichment analysis) in the clusters. The results of graph clustering with pathway enrichment analysis were presented in Table I.

Significant GO Terms (*p*-value 0.05 using hypergeometric test) were related with Neuroactive ligand-receptor interaction (hsa04080) and p53 signaling pathway (hsa04115) (Table I). EDNRB (endothelin receptor type B), ADRB2 (βadrenergic receptor), S1PR1 (sphingosine-1-phosphate receptor 1), P2RY14 (PsY purinoceptor 14), LEPR (leptin-receptor), GHR (growth hormone receptor) enriched in the Neuroactive ligand-receptor interaction (in Cluster 1 and 4) and PPM1D (protein phosphatase-1D), GADD45B (growth arrest and DNA-damage-inducible, beta) enriched in p53 signaling pathway (in Cluster 3) (not displayed in Table I).

Discussion

According to our analysis results, we could find that 1030 DEGs have been identified by our

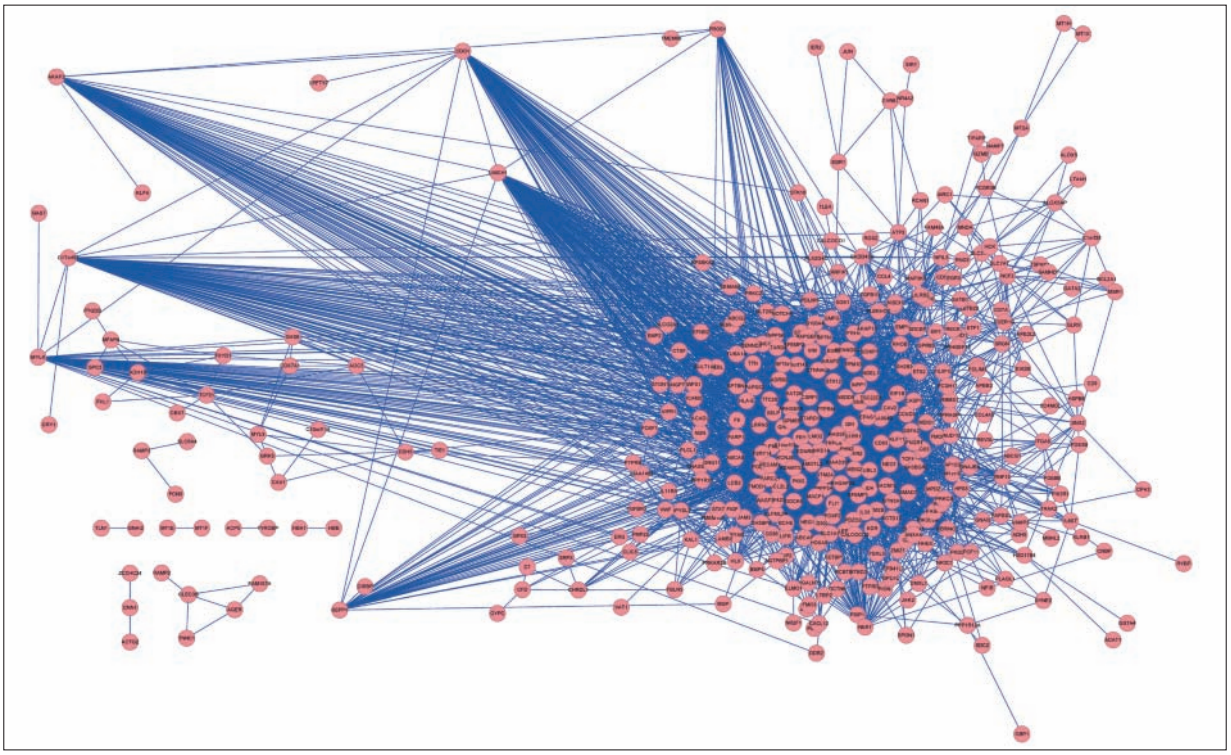

Figure 1. Correlation network of LAC. The point stands for DEG and the blue line strands for the correlation of two neighbor points with the r > 0.9.

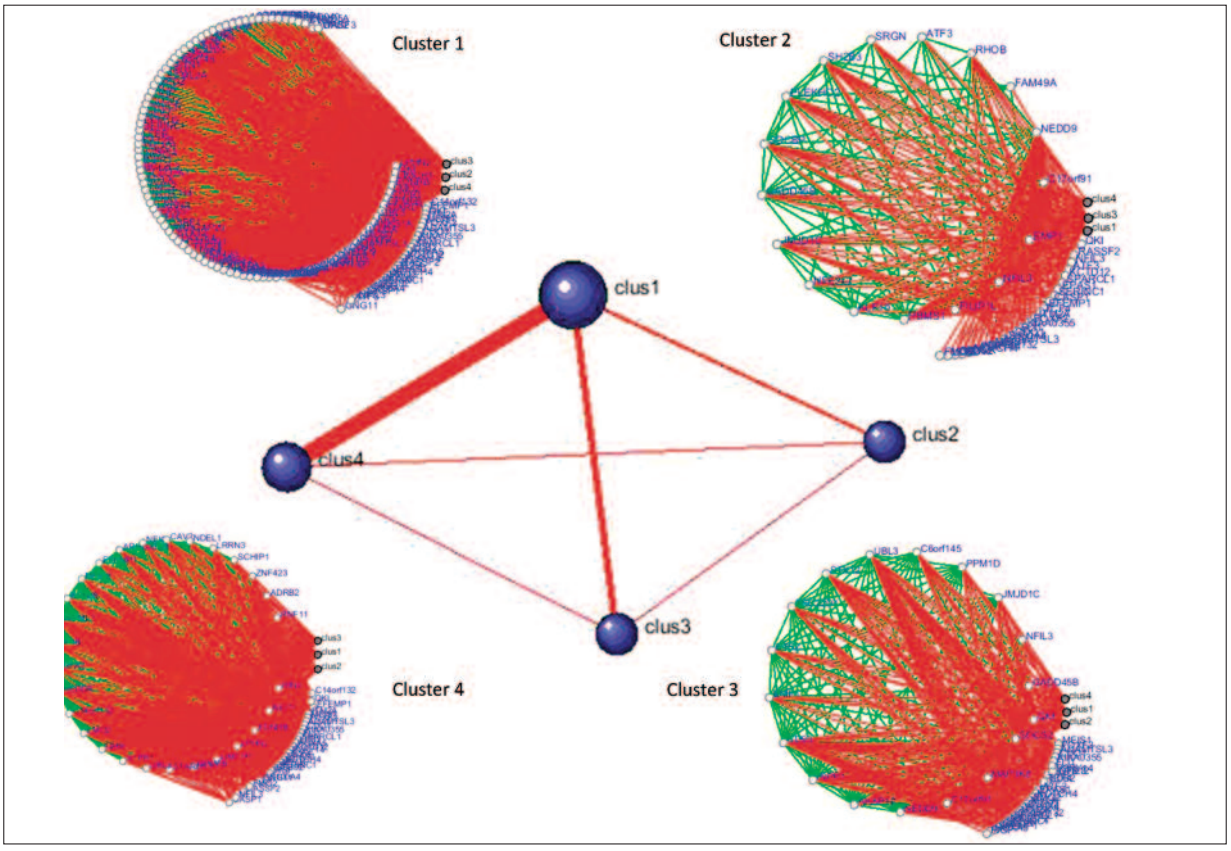

Figure 2. Graph clustering of correlated modules in LAC (threshold r ≥ 0.9).

Table I. List of enriched pathways in cluster1 to 4 detected by DPClus.

Category	Term	Description	*p*-value	FDR
Cluster 1	hsa04080	Neuroactive ligand-receptor interaction	0.020516	0.545121
Cluster 2				
Cluster 3	hsa04115	p53 signaling pathway	0.077638	0.620844
Cluster 4	hsa04080	Neuroactive ligand-receptor interaction	0.001169	0.009317

meta-analysis and a correlation network was constructed to produce 4509 relationships among 377 DEGs. Further, those 377 DEGs were clustered into four clusters through graph clustering method in response to LAC. Among them, the genes of cluster 1 and 4 may be involved in LAC development by Neuroactive ligand-receptor interaction, but the genes in cluster 3 were mainly associated with LAC progression through p53 signaling pathway. Therefore, we would discuss the relationships between LAC and some identified genes as follows based on previous reports.

EDNRB, ADRB2, S1PR1, and GHR genes belong to cluster 1 and 4. P2RY14 and LEPR were also classified into cluster 1. There was some evidence that ADRB2, S1PR1, and GHR were related with LAC by interaction with their corresponding ligands. For example, activation of type I IL-1 receptors in A549 cells leads to a cell density-dependent, selective up-regulation of ADRB2, and that the mechanism of this effect involves increased formation and stability of the ADRB2 mRNA[19]. However, further study found ADRB2 expression and its functional coupling to adenylyl cyclase was not very high in the A549 cell lines, but in Calu-3 (caucasian lung, adenocarcinoma cultured) and 16HBE14o (16 human bronchial epithelium cell line)-cell, suggesting that the cell lines Calu-3 and 16HBE14o- present suitable models to study function and regulation of the β-adrenoceptor signalling in the respiratory system[20]. In addition, significant association was observed between LAC risk and ADRB2 SNP or their combined haplotypes, including G-1023A, A46G (Gly16Arg), and the haplotype $A^{-1023}A^{46}$ in the subgroup of young subjects ≤ 50 years old[21].

Sphingosine1-phosphate (S1P), a potent lipid mediator produced from the metabolism of sphingomyelin, is involved in the regulation of a number of biological activities, such as cell migration, proliferation, cytoskeleton reorganization, etc[22,23]. Such responses require the function of S1P G protein-coupled receptors S1P1, which was originally isolated as an inducible gene from endothelial cells[24]. The sphingosine 1-phosphate (S1P) receptor-1 (S1P < 1) has also been demonstrated to involve in stabilization of nascent blood vessels that is critical for step for the LAC growth and spread[25].

Recent observations revealed that the growth hormone receptor (GHR) might play important roles in LAC development by interaction with growth hormone in human lung epithelial cells[26]. And the Thr495Pro polymorphism of *GHR* was found strongly associated with lung cancer risk in Caucasians living in the UK and Chinese people[27].

Although there was no report of EDNRB, P2RY14, and LEPR involved in LAC, they all have been demonstrated play an important role in other cancers, such as nasopharyngeal carcinoma[28], prostate adenocarcinoma[29], endometrioid endometrial adenocarcinoma[30], breast cancer[31,32]. Therefore, they may be also potential makers for LAC.

Besides, PPM1D and GADD45B were two genes that clustered into cluster 3. PPM1D (protein phosphatase-1D) is a wild-type p53-induced protein phosphatase 1 that has been established with oncogenic functions in gastric carcinoma[33], breast cancer[34], and ovarian clear cell carcinomas[35]. Current evidence suggests that PPM1D has significant value for tumor progression and the clinical prognosis of patients with primary LAC. Increased PPM1D expression was observed in 64.3% of the LAC cases. PPM1D expression was found to be correlated significantly with two clinicopathological factors: γ-H2AX expression, and invasion to the pulmonary vein. Increased PPM1D expression was significantly associated with a lower overall survival rate. The Ki67 index level was also higher in the Wip1-positive group than in the negative group[36].

GADD45B is a member of evolutionarily conserved, small, acidic, nuclear protein family, which have been implicated in terminal differentiation, growth suppression, and apoptosis through specifically interaction with the Cdk1/CyclinB1 complex. A novel synthetic retinoid 6-[3-(1-adamanty1)-4-hydroxyphenyl]-2-naphthalene carboxylic acid (AHPN) has been shown to inhibit LAC cell

growth and induce apoptosis. After treatment with AHPN, there was a rapid increase in the level of GADD45B mRNA. AHPN increases the half-life of GADD45B mRNA by 9-fold, indicating that it causes an increase in the stability of these mRNAs. In brief, the inhibition of LAC cell growth by AHPN is accompanied by an increase in GADD45B mRNA, and that this enhancement is regulated at a post-transcriptional level[37,38].

Conclusions

It is effective to identify potential molecular marker for LAC and predict their underlying functions by bioinformatics analysis and graph clustering method. However, further experiments are still indispensable to confirm our conclusion.

References

1) Hanada S, Maeshima A, Matsuno Y, Ohta T, Ohki M, Yoshida T, Hayashi Y, Yoshizawa Y, Hirohashi S, Sakamoto M. Expression profile of early lung adenocarcinoma: identification of MRP3 as a molecular marker for early progression. Am J Pathol 2008; 216: 75-82.
2) Barletta JA, Yeap BY, Chirieac LR. Prognostic significance of grading in lung adenocarcinoma. Cancer 2010; 116: 659-669.
3) Nakamura N, Kobayashi K, Nakamoto M, Kohno T, Sasaki H, Matsuno Y, Yokota J. Identification of tumor markers and differentiation markers for molecular diagnosis of lung adenocarcinoma. Oncogene 2006; 25: 4245-4255.
4) Yamaguchi E, Nakayama T, Nanashima A, Matsumoto K, Yasutake T, Sekine I, Nagayasu T. Ets-1 protooncogene as a potential predictor for poor prognosis of lung adenocarcinoma. Tohoku J Exp Med 2007; 213: 41-50.
5) DiFeo A, Feld L, Rodriguez E, Wang C, Beer DG, Martignetti JA, Narla G. A functional role for KLF6-SV1 in lung adenocarcinoma prognosis and chemotherapy response. Cancer Res 2008; 68: 965-970.
6) Wang R, Geng J, Wang J, Chu X, Geng H, Chen L. Overexpression of eukaryotic initiation factor 4E (eIF4E) and its clinical significance in lung adenocarcinoma. Lung Cancer 2009; 66: 237-244.
7) Nakai R, Maniwa Y, Tanaka Y, Nishio W, Yoshimura M, Okita Y, Ohbayashi C, Satoh N, Ogita H, Takai Y. Overexpression of NECL-5 correlates with unfavorable prognosis in patients with lung adenocarcinoma. Cancer sci 2010; 101: 1326-1330.
8) Minamiya Y, Ono T, Saito H, Takahashi N, Ito M, Mitsui M, Motoyama S, Ogawa J. Expression of histone deacetylase 1 correlates with a poor prognosis in patients with adenocarcinoma of the lung. Lung Cancer 2011; 74: 300-304.
9) Minamiya Y, Ono T, Saito H, Takahashi N, Ito M, Motoyama S, Ogawa J. Strong expression of HDAC3 correlates with a poor prognosis in patients with adenocarcinoma of the lung. Tumor Biol 2010: 31: 533-539.
10) Araki K, Osaki M, Nagahama Y, Hiramatsu T, Nakamura H, Ohgi S, Ito H. Expression of RUNX3 protein in human lung adenocarcinoma: Implications for tumor progression and prognosis. Cancer Sci 2005; 96: 227-231.
11) Nose N, Sugio K, Oyama T, Nozoe T, Uramoto H, Iwata T, Onitsuka T, Yasumoto K. Association between estrogen receptor-beta expression and epidermal growth factor receptor mutation in the postoperative prognosis of adenocarcinoma of the lung. J Clin Oncol 2009; 27: 411-417.
12) Minamiya Y, Saito H, Takahashi N, Ito M, Toda H, Ono T, Konno H, Motoyama S, Ogawa J. Expression of the chemokine receptor CCR6 correlates with a favorable prognosis in patients with adenocarcinoma of the lung. Tumor Biol 2010: 32: 197-202.
13) Minamiya Y, Saito H, Takahashi N, Ito M, Imai K, Ono T, Motoyama S, Ogawa J. Expression of the chemokine receptor CXCR4 correlates with a favorable prognosis in patients with adenocarcinoma of the lung. Lung Cancer 2010; 68: 466-471.
14) Hong F, Breitling R, McEntee CW, Wittner BS, Nemhauser JL, Chory J. Rankprod: a bioconductor package for detecting differentially expressed genes in meta-analysis. Bioinformatics 2006; 22: 2825-2827.
15) Altaf-Ul-Amin M, Shinbo Y, Mihara K, Kurokawa K, Kanaya S. Development and implementation of an algorithm for detection of protein complexes in large interaction networks. BMC Bioinformatics 2006; 7: 207.
16) Fukushima A, Kusano M, Redestig H, Arita M, Saito K. Metabolomic correlation-network modules in arabidopsis based on a graph-clustering approach. BMC Syst Biol 2011; 5: 1.
17) Kanehisa M. The kegg database. Novartis Found Symp 2002; 247: 91-101; discussion 101-103, 119-128, 244-152.
18) Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using david bioinformatics resources. Nat Protoc 2009; 4: 44-57.
19) Szentendrei T, Lazar-Wesley E, Nakane T, Virmani M, Kunos G. Selective regulation of β2-adrenergic receptor gene expression by interleukin-1 in cultured human lung tumor cells. J Cell Physiol 1992; 152: 478-485.
20) Abraham G, Kneuer C, Ehrhardt C, Honscha W, Ungemach FR. Expression of functional [beta] 2-adrenergic receptors in the lung epithelial cell lines 16HBE14o(-), Calu-3 and A549. Biochim Biophys Acta 2004; 1691: 169-179.
21) Wang H, Hao B, Chen X, Zhao N, Cheng G, Jiang Y, Liu Y, Lin C, Tan W, Lu D, Wei Q, Jin L, Lin D, He F. Beta-2 adrenergic receptor gene (ADRB2) polymorphism and risk for lung adenocarcinoma: A case-control study in a chinese population. Cancer Lett 2006; 240: 297-305.

22) Garcia JGN, Liu F, Verin AD, Birukova A, Dechert MA, Gerthoffer WT, Bamberg J, English D. Sphingosine 1-phosphate promotes endothelial cell barrier integrity by Edg-dependent cytoskeletal rearrangement. J Clin Invest 2001; 108: 689-702.

23) Vasta V, Meacci E, Catarzi S, Donati C, Farnararo M, Bruni P. Sphingosine 1-phosphate induces arachidonic acid mobilization in A549 human lung adenocarcinoma cells. Biochim Biophys Acta 2000; 1483: 154-160.

24) Kono M, Mi Y, Liu Y, Sasaki T, Allende ML, Wu YP, Yamashita T, Proia RL. The sphingosine-1-phosphate receptors S1P1, S1P2, AND S1P3 function coordinately during embryonic angiogenesis. J Biol Chem 2004; 279: 29367-29373.

25) Chae SS, Paik JH, Furneaux H, Hla T. Requirement for sphingosine 1-phosphate receptor-1 in tumor angiogenesis demonstrated by in vivo rna interference. J Clin Invest 2004; 114: 1082-1089.

26) Cheng CY, Perevedentseva E, Tu JS, Chung PH, Cheng CL, Liu KK, Chao JI, Chen PH, Chang CC. Direct and in vitro observation of growth hormone receptor molecules in a549 human lung epithelial cells by nanodiamond labeling. Appl Phys Lett 2007; 90: 163903-163903.

27) Cao G, Lu H, Feng J, Shu J, Zheng D, Hou Y. Lung cancer risk associated with Thr495Pro polymorphism of GHR in chinese population. Jpn J Clin Oncol 2008; 38: 308-316.

28) Lo KW, Tsang YS, Kwong J, To KF, Teo PML, Huang DP. Promoter hypermethylation of the EDNRB gene in nasopharyngeal carcinoma. Int J Cancer 2002; 98: 651-655.

29) Jeronimo C, Henrique R, Campos P, Oliveira J, Caballero O, Lopes C, Sidransky D. Endothelin B receptor gene hypermethylation in prostate adenocarcinoma. Am J Clin Pathol 2003; 56: 52-55.

30) Chung TKH, Cheung TH, Huen NY, Wong KWY, Lo KWK, Yim SF, Siu NSS, Wong YM, Tsang PT, Pang MW. Dysregulated microrna and their predicted targets associated with endometrioid endometrial adenocarcinoma in hong kong women. Int J Cancer 2009; 124: 1358-1365.

31) Woo HY, Park H, Ki CS, Park YL, Bae WG. Relationships among serum leptin, leptin receptor gene polymorphisms, and breast cancer in Korea. Cancer Lett 2006; 237: 137-142.

32) Cleveland RJ, Gammon MD, Long CM, Gaudet MM, Eng SM, Teitelbaum SL, Neugut AI, Santella RM. Common genetic variations in the lep and lepr genes, obesity and breast cancer incidence and survival. Breast Cancer Res Treat 2010; 120: 745-752.

33) Fuku T, Semba S, Yutori H, Yokozaki H. Increased wild-type p53-induced phosphatase 1 (wip1 or PPM1D) expression correlated with downregulation of checkpoint kinase 2 in human gastric carcinoma. Pathol Int 2007; 57: 566-571.

34) Rauta J, Alarmo EL, Kauraniemi P, Karhu R, Kuukasjarvi T, Kallioniemi A. The serine-threonine protein phosphatase PPM1D is frequently activated through amplification in aggressive primary breast tumours. Breast Cancer Res Treat 2006; 95: 257-263.

35) Tan DSP, Lambros MBK, Rayter S, Natrajan R, Vatcheva R, Gao Q, Marchiò C, Geyer FC, Savage K, Parry S, Fenwick K, Tamber N, Mackay A, Dexter T, Jameson C, McCluggage WG, Williams A, Graham A, Faratian D, El-Bahrawy M, Paige AJ, Gabra H, Gore ME, Zvelebil M, Lord CJ, Kaye SB, Ashworth A, Reis-Filho JS. PPM1D is a potential therapeutic target in ovarian clear cell carcinomas. Clin Cancer Res 2009; 15: 2269-2280.

36) Satoh N, Maniwa Y, Bermudez VP, Nishimura K, Nishio W, Yoshimura M, Okita Y, Ohbayashi C, Hurwitz J, Hayashi Y. Oncogenic phosphatase wip1 is a novel prognostic marker for lung adenocarcinoma patient survival. Cancer Sci 2011; 102: 1101-1106.

37) Vairapandi M, Balliet AG, Hoffman B, Liebermann DA. GADD45B and GADD45G are cdc2/cyclinB1 kinase inhibitors with a role in S and G2/M cell cycle checkpoints induced by genotoxic stress. J Cell Physiol 2002; 192: 327-338.

38) Sakaue M, Adachi H, Jetten AM. Post-transcriptional regulation of myd118 and gadd45 in human lung carcinoma cells during 6-[3-(1-adamantyl)-4-hydroxyphenyl]-2-naphthalene carboxylic acid-induced apoptosis. Mol Pharmacol 1999; 55: 668-676.

39) Newman D, Sakaue M, Koo JS, Kim KS, Baek SJ, Eling T, Jetten AM. Differential regulation of nonsteroidal anti-inflammatory drug-activated gene in normal human tracheobronchial epithelial and lung carcinoma cells by retinoids. Mol Pharmacol 2003; 63: 557-464.