RESEARCH COMMUNICATION

Transcriptome-wide Network Analysis of Squamous Lung Cancer Reveals Potential Methylation Genes

Jiatao Lou[1], Jian Xue[1], Qiang Lin[2*]

Abstract

Lung cancer is a prevalent cancer with a high death rate. Underlying mechanisms havebeen found to be highly correlated with epigenetics, especially with DNA methylation. With methylation array and other regulation data, we constructed a TF-gene regulation network and a TF-pathway network. Through those networks, we identified lung cancer related genes that were found by previous studies, and supposed a number of new examples. Our work demonstrated the new potential methylation for lung cancer.

Keywords: Methylation - lung cancer - microarray - pathway regulation

Asian Pacific J Cancer Prev, **12**, 2349-2352

Introduction

Lung cancer is very often a deadly disease with 5-year survival rates of about 14%. It have shown that epigenetics plays an important role in cancer biology. High throughput genome-wide methylation analysis methods have been applied in cancer research. One of these approaches was Restriction Landmark Genomic Scanning (RLGS). The frequently used assays for high throughput methylation analysis include expression microarray analysis, restriction landmark genomic scanning, BeadArray-based methylation analysis, and microarray analysis in combination with immunoprecipitation of methylated DNA . Using these high throughput approaches, methylation status of thousands of genes has been analyzed. For lung cancer, more than 40 genes were found to have some degree of alteration in DNA methylation patterns, such as DAPK, APC (4), RASSF1 .

DNA methylation contributes to the regulation of the transcriptional activity of not only one gene but also a set of genes. DNA methylation are key regulation mechanism which affects the binding of transcription factors to DNA, resulting either in gene activation or gene silencing . Numerous genes which are involved in different pathways relevant for lung cancer pathogenesis have been identified. Together with other epigenetic mechanisms, the methylation of these genes leads to gene silencing. The DNA methylation of TFs plays a vital role in the transcriptional regulation, even in pathway regulation. Many transcription regulation networks have been constructed by different methods .

In this paper, we constructed a lung cancer regulation network with methylation data. We found new pathways, genes, and regulation relationships associated with lung cancer.

Materials and Methods

Affymetrix Microarray Data

One transcription profile of squamous lung cancer GSE5816 were obtained from a public functional genomics data repository GEO (http://www.ncbi.nlm.nih.gov/geo/).

Pathway Data

KEGG (Kyoto Encyclopedia of Genes and Genomes) is a collection of online databases dealing with genomes, enzymatic pathways, and biological chemicals. The pathway database records networks of molecular interactions in the cells, and its' variants are specific to particular organisms (http://www.genome.jp/kegg/). Total 130 pathways, including 2287 genes, were collected from KEGG.

Regulation Data

There are approximately 2600 proteins in the human genome that contain DNA-binding domains, and most of which are supposed to function as transcription factors. These transcription factors are grouped into 5 super class families, based on the presence of conserved DNA-binding domains.

774 pairs of regulatory relationship between 219 transcription factors (TFs) and 265 target genes were collected from TRANSFAC . 5722 pairs of regulatory relationship between 102 transcription factors (TFs) and 2920 target genes were collected from TRED . Combined datasets, total 6328 regulatory relationships between 276 TFs and 3002 target genes were collected.

Differentially Expressed Genes Analysis

For the GSE5816 dataset, the limma method was used

*[1]Department of Laboratory Medicine, [2]Department of Thoracic Surgery, Shanghai Chest Hospital Affiliated to Shanghai Jiaotong University, Shanghai ,China *For correspondence: qianglindr@gmail.com*

to identify DEGs. The original expression datasets from all conditions were processed into expression estimates using the RMA method with the default settings implemented in Bioconductor, and then construct the linear model. The 133 DEGs with the fold change value larger than 4 were selected . Of the 133 DEGs, 67 genes with CpG islands were kept for further analysis.

Gene Ontology Analysis

The BiNGO analysis was used to identify over-represented GO categories in biological process.

Regulation Network Construction

Using the regulation data that have been collected from TRANSFAC database and TRED database, we matched the coexpression between differentially expressed TFs and its differentially expressed target genes.

Base on the above two regulation datasets and the pathway relationships of the target genes, we build the regulation networks by Cytoscape . Base on the significant relationships (PCC > 0.6 or PCC < -0.6) between TFs and its target genes, 57 putative regulatory relationships were predicted between 12 TFs and 45 target genes.

Significance Analysis of Pathway

We adopted an impact analysis that includes the statistical significance of the set of pathway genes but also considers other crucial factors such as the magnitude of each gene's expression change, the topology of the signaling pathway, their interactions, etc al. In this model, the Impact Factor (IF) of a pathway Pi is calculated as the sum of two terms:

$$IF(Pi) = \log(\frac{1}{pi}) + \frac{\sum_{g \in Pi} |PF(g)|}{|\Delta E| \cdot N_{de}(Pi)}$$

The first term is a probabilistic term that captures the significance of the given pathway Pi from the perspective of the set of genes contained in it.

It is obtained by using the hyper geometric model in which pi is the probability of obtaining at least the observed number of differentially expressed gene, Nde, just by chance.

The second term is a functional term that depends on the identity of the specific genes that are differentially expressed as well as on the interactions described by the pathway.

The second term sums up the absolute values of the perturbation factors (PFs) for all genes g on the given pathway Pi.

The PF of a gene g is calculated as follows:

$$PF(g) = \Delta E(g) + \sum_{u \in USg} \beta_{ug} \cdot \frac{PF(u)}{N_{ds}(u)}$$

In this equation, the first term ΔE (g) captures the quantitative information measured in the gene expression experiment. The factor ΔE (g) represents the normalized measured expression change of the gene g. The first term ΔE (g)in the above equation is a sum of all PFs of the genes u directly upstream of the target gene g, normalized by the number of downstream genes of each such gene Nds(u), and weighted by a factor βug, which reflects the type of interaction: βug = 1 for induction, βug = –1 for repression (KEGG supply this information about the type of interaction of two genes in the description of the pathway topology). USg is the set of all such genes upstream of g. We need to normalize with respect to the size of the pathway by dividing the total perturbation by the number of differentially expressed genes on the given pathway, Nde(Pi). In order to make the IFs as independent as possible from the technology, and also comparable between problems, we also divide the second term in equation 1 by the mean absolute fold change ΔE, calculated across all differentially expressed genes.

Regulation Network between TFs and Pathways

To further investigate the regulatory relationships between TFs and pathways, we mapped DEGs to pathways and got a regulation network between TFs and pathways.

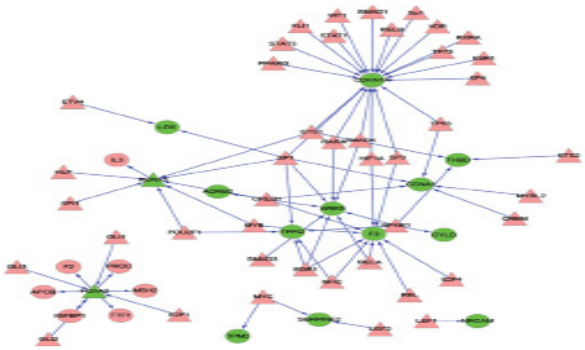

Figure 1. Regulation Network of Squamous Lung Cancer. Triangles denote transcription factors and circles targeting genes. Green nodes stand for methylation genes and pink for other genes.

Results

Regulation Network Construction

To get DEGs of lung cancer methylation, we obtained publicly available microarray data set pmid0030486 . After normalization analysis of pmid0030486 dataset, the 67 differentially expressed genes with the fold change value larger than 4 were selected. After mapped to the regulation datasets (TRANSFAC and TRED database), we got 77 regulatory relationships between 44 TFs and their 21 target genes. By integrating the regulatory relationships above, a regulation network of lung cancer methylation was built between TFs and its target genes (Figure 1). In this network, CDKN1A as a methylation gene regulates lots of target genes with a higher degree form a local network, suggesting roles in lung cancer.

GO Analysis of he Regulation Network in SCC

Several Gene Ontology (GO) categories were enriched among these genes in the regulatory network, including positive regulation of biological process, positive regulation of cellular process and positive regulation of macromolecule biosynthetic process (Figure 2).

Regulation Network between TFs and Pathways

To further investigate the regulatory relationships

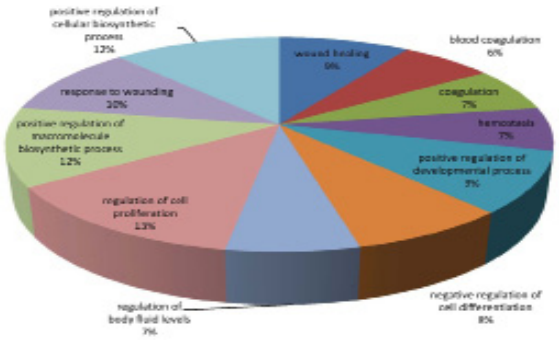

Figure 2. GO-biological Process Analysis. The pie chart listd the significant GO categories with the p-value <0.05.

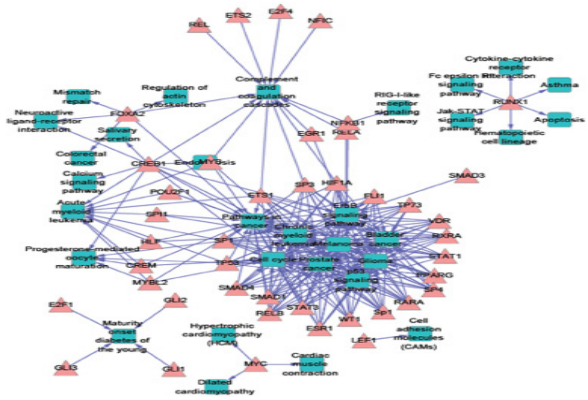

Figure 3. Regulation Network of the TF-PATHWAY. Triangles denote transcription factor, and rectangles denote pathways.

between TFs and pathways, we mapped methylation genes to pathways and constructed a regulation network between TFs and pathways (Figure 3). In the network, acute myeloid leukemia, complement and coagulation cascades, pathways in cancer and melanoma pathways shown as hub nodes were regulated by many TFs.

Discussion

lLung cancer accounts for the most cancer related deaths for both men and women. DNA methylation in the promoter region of certain genes is associated with transcriptional silencing. Methylation affects gene expression directly by interfering with transcription factor binding and/or indirectly by recruiting histone deacetylases from methyl-DNA-binding proteins. The identification of methylation associated genes in lung cancer is essential to provide an important possibility in the prevention of this disease. In our method, we used public functional genomic data and regulation data to construct a network between TFs and their target genes. From the constructed regulation network in lung cancer, many TFs and pathways, which are highly related to lung cancer, have been linked by our method. The gene CDKN1A, RUNX1, TFPI2, AREG, FOXA2 are also hub nodes in our transcriptome network and have a close relationship with lung cancer proved by previous studies. Although the role of F3 in lung cancer has not been investigated to date, some evidence suggests that F3 may play an important role in lung cancer.

CDKN1A, also known as P21, encodes a potent cyclin-dependent kinase inhibitor. The encoded protein binds to and inhibits the activity of cyclin-CDK2 or -CDK4 complexes, and thus functions as a regulator of cell cycle progression at G1. It have been suggested that hypermethylation around the consensus Sp1-binding sites may directly reduce Sp1/Sp3 binding, therefore leading to a reduced p21(Cip1) expression in response to depsipeptide treatment .

RUNX1 is susceptible to aberrant methylation and associated with lung cancer as a DNA marker . RUNX1 has been shown to interact with SUV39H1, a histone methyltransferase responsible for histone H3K9 methylation, and this interaction disrupts the binding of RUNX1 to DNA .

CCNA1 were previously described to be frequently methylated in cervical cancer .

The role of AREG in cancer development and progression is also supported by clinical data. It has been established a significant correlation between elevated AR mRNA levels in bladder tumour tissue and poor patient survival. In patients with advanced non-squamous non-small cell lung cancers increased levels of circulating AR in serum are predictors of poor response to Gefitinib, which is a drug used in the treatment of certain types of cancer.

Tissue factor pathway inhibitor-2 (TFPI-2) is a Kunitz-type serine proteinase inhibitor that inhibits plasmin-dependent activation of several metalloproteinases. Down regulation of TFPI-2 could thus enhance the invasive potential of neoplastic cells in several cancers, including lung cancer. In non-small-cell lung cancer, TFPI-2 promoter hypermethylation was frequently found in patients with late-stage cancer (stages III and IV) and with lymph node metastases.

Down-regulation of HNF3 beta expression in a large proportion of lung cancer cell lines examined and identified two novel mutants of HNF3 beta, as well as hypermethylation of the HNF3 beta promoter .Conditional expression of HNF3 beta led to significant growth reduction, proliferation arrest, apoptosis, and loss of clonogenic ability, suggesting additionally that HNF3 beta is a novel tumor suppressor in lung cancer. It have been showed that genetic abnormalities of lung-specific differentiation pathways in the development of lung cancer. The protein encoded by F3 is a member of the immunoglobulin superfamily. It is a glycosylphosphatidylinositol (GPI)-anchored neuronal membrane protein that functions as a cell adhesion molecule. It may play a role in the formation of axon connections in the developing nervous system. Two alternatively spliced transcript variants encoding different isoforms have been described for this gene.

This gene encodes coagulation factor III which is a cell surface glycoprotein. This factor enables cells to initiate the blood coagulation cascades, and it functions as the high-affinity receptor for the coagulation factor VII. The resulting complex provides a catalytic event that is responsible for initiation of the coagulation protease cascades by specific limited proteolysis. Unlike the other cofactors of these protease cascades, which circulate as nonfunctional precursors, this factor is a potent initiator that is fully functional when expressed on cell surfaces. This protein is the only one in the coagulation pathway

for which a congenital deficiency has not been described.

TP53 and its downstream target gene, CDKN1A were demonstrated to play an important role in sagopilone (SAG) chemotherapeutic agent sensitivity in lung cancer. Treatment of A549 cells with a low concentration of SAG revealed an up-regulation of CDKN1A. TP53 mediates G1 arrest mainly by increasing protein levels of CDKN1A. Furthermore, knockdown of TP53, which inhibited the transcriptional induction of CDKN1A, led to a significant increase in apoptosis induction in A549 cells when treated with a low concentration of SAG. These results indicated that activation of TP53 and its downstream effectors CDKN1A by low concentrations of SAG is responsible for the relative apoptosis resistance of A549 cells .

The PPARG is a member of the nuclear receptor superfamily of ligand-dependent transcriptional factors. Previous reports indicated that PPARG ligands inhibit human lung cancer cell growth and induce apoptosis by stimulating the cyclin-dependent kinase inhibitor CDKN1A and by reducing cyclin D1 gene expression. The induction of CDKN1A gene expression by PPARG ligands may be mediated through increased Sp1- and NF-IL6 (C/EBP)-dependent transcriptional activation, two transcription factors with regulatory elements in the promoter region of the CDKN1A gene. These observations unveil a mechanism for CDKN1A gene regulation in lung cancer that represents a potential target for therapy .

Tamoxifen(Tam) was reported to exert an anti-cancer effect on a number of estrogen receptor (ER)-negative lung cancer. Inhibition of growth of ER-negative human lung cancer cells by Tam is associated with the induction of CDKN1A, and mutation of Sp1-binding sites dramatically attenuated Tam-induced CDKN1A promoter activity. Furthermore, PKA activators activated CDKN1A promoter activity and increased CDKN1A protein level in lung cancer cells. Taken together, these results demonstrated that Tam activated the CDKN1A promoter via Sp1-binding sites and suggested that PKA may be involved in the induction of CDKN1A by Tam in ER-negative lung cancer cells .

The IF analysis method yields 11 pathways significant at the 5% level and the top five pathways contained complement and coagulation cascades: ECM-receptor interaction, P53 signaling pathway, cell adhesion molecules (CAMs), Focal adhesion, Cell cycle.

The impact analysis agrees well with known lung cancer related pathways which are supported by the existing literature, such as P53 signaling pathway, cell adhesion molecules (CAMs), focal adhesion and cell cycle. Moreover, this analysis also identifies additional lung cancer related pathways that have not been detected, such as complement and coagulation cascades.

In this research, we identified pathways crosstalk network by integrating biological pathways and expression data. Many expected pathway crosstalk have been identified by this approach. Sets of key pathway were identified that can be found in the lung cancer disease. It turned out that this analysis method is well suited for microarray data and therefore is proposed as a powerful tool for the search for new and so far undiscovered pathways related to other cancer.

References

Agathanggelou A, Honorio S, Macartney DP, et al (2001). Methylation associated inactivation of RASSF1A from region 3p21.3 in lung, breast and ovarian tumours. *Oncogene*, **20**, 1509-18.

Feng Q, Hawes SE, Stern JE, et al (2008). DNA methylation in tumor and matched normal tissues from non-small cell lung cancer patients. *Cancer Epidemiol Biomarkers Prev*, **17**, 645-54.

Gu F, Hsu HK, Hsu PY, et al (2010). Inference of hierarchical regulatory network of estrogen-dependent breast cancer through ChIP-based data. *BMC Syst Biol*, **4**, 170.

Han S, Sidell N, Fisher PB, Roman J (2004). Up-regulation of p21 gene expression by peroxisome proliferator-activated receptor gamma in human lung carcinoma cells. *Clin Cancer Res*, **10**, 1911-9.

Harden SV, Tokumaru Y, Westra WH, et al (2003). Gene promoter hypermethylation in tumors and lymph nodes of stage I lung cancer patients. *Clin Cancer Res*, **9**, 1370-5.

Jiang C, Xuan Z, Zhao F, Zhang MQ (2007). TRED: a transcriptional regulatory element database, new entries and other development. *Nucleic Acids Res*, **35**(Database issue), D137-140.

Jones PA (2002), Baylin SB. The fundamental role of epigenetic events in cancer. *Nat Rev Genet*, **3**, 415-28.

Kanehisa M (2002). The KEGG database. Novartis Foundation symposium;247:91-101; discussion 101-3, 119-28, 244-52.

Kawakami K, Brabender J, Lord RV, et al (2000). Hypermethylated APC DNA in plasma and prognosis of patients with esophageal adenocarcinoma. *JNCI*, **92**, 1805-11.

Kitkumthorn N, Yanatatsanajit P, Kiatpongsan S, et al (2006). Cyclin A1 promoter hypermethylation in HPV-associated cervical cancer. *BMC Cancer*, **6**, 55.

Lee TH, Chuang LY, Hung WC (200). Induction of p21WAF1 expression via Sp1-binding sites by tamoxifen in estrogen receptor-negative lung cancer cells. *Oncogene*, **19**, 3766-73.

Maere S, Heymans K, Kuiper M (2005). BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. *Bioinformatics*, **21**, 3448-9.

Matys V, Fricke E, Geffers R, et al (2003). TRANSFAC: transcriptional regulation, from patterns to profiles. *Nucleic Acids Res*, **31**, 374-8.

Seki M, Umezawa T, Urano K, Shinozaki K (2007). Regulatory metabolic networks in drought stress responses. *Curr Opin Plant Biol*, **10**, 296-302.

Shames DS, Girard L, Gao B, et al (2006). A genome-wide screen for promoter methylation in lung cancer identifies novel methylation markers for multiple malignancies. *PLoS Med*, **3**, e486.

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

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

Weber M, Davies JJ, Wittig D, et al (2005). Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells. *Nat Gene*, **37**, 853-62.

Winsel S, Sommer A, Eschenbrenner J, et al (2011). Molecular mode of action and role of TP53 in the sensitivity to the novel epothilone sagopilone (ZK-EPO) in A549 non-small cell lung cancer cells. *PLoS One*, **6**, e19273.

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