Pathway cross-talk network strategy reveals key pathways in non-small cell lung cancer

Meng-Li Zheng¹, Nai-Kang Zhou², De-Liang Huang³, Cheng-Hua Luo⁴
¹Department of Thoracic Surgery, The 309th Hospital, PLA, Beijing 100091, China; ²Department of Thoracic Surgery, General Hospital, PLA, Beijing 100855, China; ³Department of Oncology, Modern Hospital Guangzhou, Guangzhou 510000, China; ⁴Department of General Surgery, Peking University International Hospital, Beijing 100026, China

Summary

Purpose: The purpose of this study was to explore the pathway cross-talks and key pathways in non-small cell lung cancer (NSCLC) to better understand the underlying pathological mechanism.

Methods: Integrated gene expression data, pathway data and protein-protein interaction (PPI) data were assessed to identify the pathway regulatory interactions in NSCLC, and constructed the background and disease pathway cross-talk networks, respectively. In this work, the attractor method was implemented to identified the differential pathways, and the rank product (RP) algorithm was used to determine the importance of pathways.

Results: Based on 787,896 PPI interactions from STRING database and 300 human pathways from KEGG, we constructed the back pathway cross-talk network with 300 nodes and 42239 edges. Integrating with expression data of NSCLC, each pathway cross-talk endowed with a weight value, and disease pathway cross-talks were identified. By RP algorithm and topology analysis of network, we selected 5 key pathways, including Alanine, DNA replication, Fanconi anemia pathway, Cell cycle and MicroRNAs in cancer under the pre-set thresholds.

Conclusion: We successfully revealed the disease pathway cross-talks and explored 5 key pathways in NSCLC, which may be the underlying therapeutic targets for lung cancer.

Key words: non-small cell lung cancer, pathway crosstalk network, protein-protein interactions, rank product

Introduction

Lung cancer is the most common cancer and the leading cause of cancer-related death around the world, resulting in more than one million deaths annually [1]. In contrast to the advances in survival for most cancers, lung cancer presents a relative poor prognosis with the 5-year survival of 18%, partially due to late-stage detection and inefficient late-stage treatments [2]. Approximately, 80% of patients with lung cancer have NSCLC [3]. Currently, treatment of NSCLC is mainly based on histopathologic characteristics and disease stage, while patients with similar pathological features and comparable stages usually present various responses to the same treatment. Numerous reports have indicated that the combination of environmental exposure and genetic susceptibility contribute to oncogenesis [4-8]. It is predicted that molecular therapeutics might be an effective approach for the prevention and treatment of lung cancer.

Current advances in high-throughput technologies enable investigators to reveal the molecular genetic characteristics of lung cancer [9-11]. Numerous studies have reported very clear influences of individual genes [12], functional pathways [13] and PPIs [14] on the development of disease. In particular, biological pathways in human cell function together in a highly orches-
Pathway cross-talks in non-small cell lung cancer

In biological systems, pathways function together in a highly coordinated manner in order to respond appropriately to various stimuli. Herein we assumed that if more PPIs could be detected between two pathways than expected by chance, these two pathways were inclined to interact with each other, that is to say, these two pathways had a cross-talk. In order to reveal the functions of pathways in cells, we firstly investigated the cross-talks of all background pathways. Based on this hypothesis, pathway interactions were explored by systematically combining both pathway data and PPI network data. A total of 300 KEGG human pathways (covering 6919 genes) were recruited in this work. In a given pathway, if a gene had interactions, we counted the number of interactions and randomly drew a gene form PPI data which interacted with the same number of genes, and then replaced the original one with the randomly selected one. Once both pathways were randomized, a cross-talk was produced. Combining both pathway data and PPI network data mentioned above, we calculated background distribution of protein interaction count number of each pathway pair, and constructed the background pathway cross-talk network, in which the interaction count of each pathway pair was defined as the weight value of a cross-talk. In our study, this randomization was repeated 10000 times. The detailed procedure was similar to a previous study of Li et al. [22].

Disease pathway cross-talks

To investigate the disease pathway cross-talks, two steps were implemented: firstly, we identified the differential pathways in NSCLC using the attract method; secondly, we constructed disease pathway cross-talks based on differential pathways.

Differential pathways

The attract method, a knowledge-driven analytical approach for identifying and annotating the gene-sets, was applied to explore differential pathways in NSCLC [23]. To facilitate a screening process, two steps were implemented to narrow the number of pathways to a reasonable number of final candidates. One was to intersect the genes in the KEGG pathways with gene expression profile of NSCLC and retain the common genes for further analysis; the other was to remove the pathways with less than 5 genes or more than 100 genes for further analysis; the other was to remove the pathways with less than 5 genes or more than 100 genes.
genes, because pathways with too many genes might be too generic and pathways with too few genes may not have sufficient biological content [22]. In detail, the KEGG pathways based on genes in expression profile were obtained based on the Database for Annotation, Visualization and Integrated Discovery (DAVID) [24]. A Fisher’s (F) exact test was performed to identify the core pathways, for gene i, \( F^0(i) \) was computed:

\[
F^0(i) = \frac{1}{K-1} \sum_{k=1}^{K} r_i \left[ y_i^{(k)} - y_i^{(0)} \right]^2
\]

where \( j \) represented the corresponding expression value in each replicate sample; \( n \) for each cell type \( k=1, \ldots, K \), \( y \) stood for the mixed effect model; \( N \) meant the total number of samples. Large values of the \( F \)-statistic indicated a strong association whereas a small \( F \)-statistic suggested that the gene demonstrated minimal cell type-specific expression changes. To make the \( F \)-statistic more confident, we selected \( T \) test to correct the log2-transformed \( F \)-statistic and obtain \( p \) value for each pathway. Adjusting the \( p \) values on the basis of false discovery rate (FDR) [25], we arranged the pathways in descending order of \( p \) values. In this article, we defined the pathways as differential pathways under the \( p \) value < 0.05.

**Disease pathway cross-talks**

Unlike background pathway cross-talk network, the disease pathway cross-talk network was based on the differential pathways we identified. The screen of cross-talks among differential pathways was similar to that of background pathways. To identify the disease pathway cross-talks, we calculated the weight value for each cross-talk. For each cross-talk, its weight value was defined as the total absolute different value of Spearman correlation coefficient (SCC) between NSCLC and normal controls. To determine the threshold value, the null hypothesis was that the ratio of true interactions between two pathways is 1, for all possible interactions was the same as the ratio of random interactions to all random interactions. We randomly extracted the genes number of G1, G2 from the expression profile, where G1, G2 represented the length of two pathways, respectively. Moreover, we performed the PPI analysis for the G1, G2 genes. If they had interactions, we calculated the weight value of these two pathways. If not, we defined it as zero. A randomized analysis of 10000 times was performed, and the corresponding 10000 weight values were obtained. We adjusted the weight value using Benjamin-Hochberg FDR-based method [25] to obtain the adjusted \( p \) value. In this work, we only focused on cross-talks with adjusted \( p < 0.01 \), and constructed the disease pathway cross-talk network.

**Key pathways**

In order to gain the critical tumorigenic pathways, rank product (RP) algorithm [26], a simple yet powerful meta-analysis tool to detect differentially expressed genes between two experimental conditions, was implemented to perform analysis on the two networks. In this work, let \( U \) and \( V \) stand for two conditions (NSCLC vs controls), and there were \( n_u \) and \( n_v \) replicates in background pathway cross-talk network, \( m_u \) and \( m_v \) in disease pathway cross-talk network. The RP for each cross-talk was determined according to the following formula:

\[
RP_s = \left( \prod_{i=1}^{n_s} r_{si} \right)^{1/T}
\]

where \( r_{si} \) stood for the rank of \( s \)th gene under \( i \)th comparison, \( i = 1, \ldots, T \). The pathways of whose \( RP \) value was < 0.01 were considered as significant pathways.

Meanwhile, we conducted the topology analysis [27] for two pathway cross-talk networks to deeply investigate the significance of hub cross-talks. The topology analysis mainly contains degree, closeness, betweenness and transitivity, in which degree quantifies the local topology of each node by summing up the number of its adjacent nodes [28]. It gives a simple count of the number of interactions of a given node and is particularly useful to identify key players in biological processes.

Additionally, the impact factor (IF) was considered to determine the importance of pathways. For an arbitrary pathway \( x \), \( p_x \) represented the degree value of pathway \( x \) in disease pathway cross-talk network, and \( p_x \) represented the \( p \) value according to the Attract method. The IF of pathway \( x \) was calculated according to the following formula:

\[
IF_x = \frac{p_x}{1 - p_x}
\]

In the present study, the pathways under \( p_A < 0.01 \), RP value < 0.01, as well as IF > 140 were considered as hub pathway.

**Results**

**Background pathway cross-talk network**

In this study, we collected a total of 787,896 PPI interactions (covering 16,730 genes) from STRING database and 300 human pathways (covering 6919 genes) from KEGG database. Combining pathway data with PPI data, a total of 42239 cross-talks were produced among 300 background pathways, as shown in Figure 1. The background pathway cross-talk network was regarded as a graph in which nodes were pathways and edges represented the cross-talk between pathways. In this network, there were great overlaps among any two cross-talks, indicating the small difference between NSCLC and normal controls. The total degree distribution also gave proof for the great overlaps (Figure 2). Moreover, edges between two pathways with significant gene overlap were considered as not informative and were removed from the network.
Disease pathway cross-talk network

In order to discover cross-talks among different biological activities in NSCLC, disease pathway cross-talk network was constructed. After removing pathways with the number of genes < 5 and > 100, a total of 283 pathways remained. Pathways under the adjusted p value < 0.01 were used to generate the disease pathway cross-talk network (Figure 3). The degree distribution of pathways in disease pathway cross-talk network was different from that in the background pathway cross-talk network (Figure 2). In disease network, the degree was decentralized relative to that in the background network. The density was highest during the degree around 200 and 300 in the background and the disease pathway cross-talk networks, respectively. This might give a hand for exploring different cross-talks between lung cancer and normal controls.

Key pathways

In order to identify significant cross-talks in pathway networks, RP algorithm in R was implemented. Under the threshold of RP < 0.01, a total of 15 significant pathways were detected. After the topology analysis for two networks, the degree of each pathway was obtained. Under $p_A < 0.01$, RP value < 0.01, as well as IF > 140, a total of 5 key pathways were identified, including Alanine (IF=187), DNA replication (IF=220), Fanconi anemia pathway (IF=144), Cell cycle (IF=204) and MicroRNAs in cancer (IF=203). The results are shown in Table 1. These key pathways affected many other pathways and showed more important roles in the process of disease. Furthermore, the key cross-talks for key pathways were identified, as illustrated in Figure 4.

<table>
<thead>
<tr>
<th>Pathway</th>
<th>p value of attractor</th>
<th>RP value</th>
<th>IF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alanine</td>
<td>0.0001</td>
<td>0.0016</td>
<td>187</td>
</tr>
<tr>
<td>DNA replication</td>
<td>0.0083</td>
<td>0.0032</td>
<td>220</td>
</tr>
<tr>
<td>Fanconi anemia pathway</td>
<td>0.0001</td>
<td>0.0049</td>
<td>144</td>
</tr>
<tr>
<td>Cell cycle</td>
<td>0.0023</td>
<td>0.0029</td>
<td>204</td>
</tr>
<tr>
<td>MicroRNAs in cancer</td>
<td>0.0050</td>
<td>0.0068</td>
<td>203</td>
</tr>
</tbody>
</table>

RP: rank product, IF: impact factor

Figure 1. Background pathway cross-talk network for non-small cell lung cancer. Nodes stood for background pathways, edges were the cross-talks among pathways. The yellow nodes represented the key pathways.

Figure 2. Total degree distribution for background pathway cross-talk network and disease pathway cross-talk network. The red line represents background condition, the blue line represents disease condition.

Figure 3. Disease pathway cross-talk network for non-small cell lung cancer. Nodes stood for background pathways, edges were the cross-talks among pathways. Yellow nodes represent the key pathways.

Table 1. Key pathways in non-small cell lung cancer
**Discussion**

NSCLC is a major public health problem worldwide and causes a great number of cancer-related deaths. There is therefore an urgent need to develop novel treatments that do not depend on conventional pharmacological approaches. Biological pathway regulation is complex, yet it underlies the functional coordination in a cell. Cancer is a disease that is characterized by unregulated growth, driven by underlying pathway deregulation. This pathway deregulation is both within pathways and between pathways. In this study we detected the functional pathways using pathway cross-talk analysis. This method has an advantage that it focuses on the collective behavior of pathways instead of individual pathways. In our study, 5 key pathways were selected based on $p_A<0.01$, RP value <0.01, as well as IF>140, including Alanine, DNA replication, Fanconi anemia pathway, Cell cycle and MicroRNAs in cancer. These findings can also be used to improve the efficacy evaluation of specific interventions on lung cancer therapy.

Cell cycle is a series of coordinated procedures, which exerts important roles of integrating the environment signal pathways with cell proliferation and cell growth [29]. It is a vital process by which a single-celled fertilized egg develops into a mature organism, as well as the process by which hair, skin, blood cells, and some internal organs are renewed [30]. The events of the cell cycle of most organisms are ordered into dependent pathways in which the initiation of late events is dependent on the completion of early events [31]. A deregulation of the cell cycle components may lead to tumor formation [32,33]. Relevant researches have reported that cell cycle regulates the liver cancer [34,35]. Moreover, recent studies also showed that cell cycle was related to the development and progress of lung cancer [36,37]. We further showed proof-of-concept evidence that the cell cycle pathway was a crucial biomarker in lung cancer in our study.

DNA replication is the process of producing two identical replicas from one original DNA molecule. This biological process occurs in all living organisms and is the basis for biological inheritance. At present, DNA replication was shown to be related to a variety of diseases. For example, defects in mitochondrial DNA replication can cause mitochondrial genetic diseases in humans due to mitochondrial DNA deletions, point mutations, or depletion which ultimately cause loss of oxidative phosphorylation [38]. DNA replication stress can be a key element of the pathogenetic cascade explaining the interplay between ectopic cell cycle events and genetic instabilities in the Alzheimer’s brain [39]. The regulation of DNA replication can influence the development of breast cancer [40]. Moreover, a recent study has pointed out that the change of DNA replication was related to lung cancer [41], which was consistent with our results.

Moreover, the other three key pathways were also reported to be associated with the progression of lung cancer. The Fanconi anemia pathway plays essential roles in response to DNA damage in cancer cell lines, and deficiencies in Fanconi anemia pathway could be considered as a predictor for personalized therapeutic treatment in patients with lung cancer [42-44]. MicroRNAs regulate target gene expression through translation repression or mRNA degradation. In NSCLC, many miRNAs have been proven to regulate the process of oncogenesis, such as miR-21 [45], miR-1254 [46]. This study demonstrated the presence of major alterations in the pathway of hepatic gluconeogenesis in weight-losing lung cancer patients. Moreover, previous studies also indicated that aberrant alanine metabolism was involved in the weight loss of lung cancer patients [47]. Taken together, these observations supported the conclusion that these hub pathways provided the targets to prevent and treat NSCLC.

In summary, we successfully revealed the disease pathway cross-talks and identified 5 key pathways in NSCLC, which might be responsible for the development and progress of NSCLC. Targeting these cross-talks may be an important new strategy for overcoming lung cancer.

**Conflict of interests**

The authors declare no conflict of interests.
References


