

## Research Article

# Identification of Lung-Cancer-Related Genes with the Shortest Path Approach in a Protein-Protein Interaction Network

Bi-Qing Li,<sup>1,2</sup> Jin You,<sup>1,3</sup> Lei Chen,<sup>4</sup> Jian Zhang,<sup>5</sup> Ning Zhang,<sup>6</sup> Hai-Peng Li,<sup>7</sup> Tao Huang,<sup>8</sup> Xiang-Yin Kong,<sup>1,3</sup> and Yu-Dong Cai<sup>9</sup>

<sup>1</sup> The Key Laboratory of Stem Cell Biology, Institute of Health Sciences, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 200031, China

<sup>2</sup> Key Laboratory of Systems Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 200031, China

<sup>3</sup> Shanghai Jiao Tong University School of Medicine, Shanghai 200025, China

<sup>4</sup> College of Information Engineering, Shanghai Maritime University, Shanghai 201306, China

<sup>5</sup> Department of Ophthalmology, Shanghai First People's Hospital, Shanghai Jiaotong University, Shanghai 200080, China

<sup>6</sup> Department of Biomedical Engineering Tianjin University, Tianjin Key Lab of BME Measurement, Tianjin 300072, China

<sup>7</sup> CAS-MPG Partner Institute for Computational Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 200031, China

<sup>8</sup> Department of Genetics and Genomic Sciences, Mount Sinai School of Medicine, New York, NY 10029, USA

<sup>9</sup> Institute of Systems Biology, Shanghai University, Shanghai 200444, China

Correspondence should be addressed to Xiang-Yin Kong; [xykong@sibs.ac.cn](mailto:xykong@sibs.ac.cn) and Yu-Dong Cai; [cai-yud@yahoo.com.cn](mailto:cai-yud@yahoo.com.cn)

Received 12 March 2013; Revised 19 April 2013; Accepted 29 April 2013

Academic Editor: Bing Niu

Copyright © 2013 Bi-Qing Li et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Lung cancer is one of the leading causes of cancer mortality worldwide. The main types of lung cancer are small cell lung cancer (SCLC) and nonsmall cell lung cancer (NSCLC). In this work, a computational method was proposed for identifying lung-cancer-related genes with a shortest path approach in a protein-protein interaction (PPI) network. Based on the PPI data from STRING, a weighted PPI network was constructed. 54 NSCLC- and 84 SCLC-related genes were retrieved from associated KEGG pathways. Then the shortest paths between each pair of these 54 NSCLC genes and 84 SCLC genes were obtained with Dijkstra's algorithm. Finally, all the genes on the shortest paths were extracted, and 25 and 38 shortest genes with a permutation  $P$  value less than 0.05 for NSCLC and SCLC were selected for further analysis. Some of the shortest path genes have been reported to be related to lung cancer. Intriguingly, the candidate genes we identified from the PPI network contained more cancer genes than those identified from the gene expression profiles. Furthermore, these genes possessed more functional similarity with the known cancer genes than those identified from the gene expression profiles. This study proved the efficiency of the proposed method and showed promising results.

## 1. Introduction

Lung cancer is one of the leading causes of cancer mortality worldwide [1]. Two main types of lung cancer are non-small cell lung cancer (NSCLC), which accounts for 80%–85%, and small cell lung cancer (SCLC), which accounts for around 20% of all cases. However, the SCLC has an extraordinarily high degree of metastasis and a strong association with smoking [2]. Diagnosis and treatment at the early stage of the disease process could reduce fatalities and increase the probability of disease-free survival. Therefore, it is meaningful

to screen lung-cancer-related genes that could be used as prognostic factors or to help elucidate the mechanism of the disease.

Recently, as high-throughput biotechnologies develop rapidly, numerous biological data have been generated from processes such as protein complex, yeast two-hybrid systems, and gene expression profiles. These data are useful resources for understanding and deducing gene function. So far, protein-protein interaction (PPI) data has been widely utilized to annotate and predict the gene function assuming that interaction proteins possess the similar or identical

functions and thus may participate in the same pathways. This so-called “guilt by association” rule was initially proposed by Nabieva et al. [3]. This rule could also be utilized to identify novel cancer-related genes.

Search Tool for the Retrieval of Interacting Genes (STRING) is an online database resource [4] that provides both predicted and experimental interaction information with a confidence score. It has been shown that proteins with short distances between each other in the PPI network tend to have the same biological functions [5–8], and interactive neighbors are prone to have the same biological functions as noninteractive ones [9, 10]. The possible reason is that the query protein and its interactive proteins might form a protein complex to exert a particular function or might participate in the same pathways.

Though great successes have been achieved for gene function prediction and identification of novel cancer-related genes with the application of the high-throughput data, yet high-throughput data is not error free. In this work, we proposed a computational method for identifying lung-cancer-related genes based on PPI network constructed from STRING. 54 NSCLC and 84 SCLC related genes were retrieved from associated KEGG pathways. Then, Dijkstra’s algorithm [11] was employed to obtain the shortest paths between each pair of the 54 NSCLC and 84 SCLC genes. All the genes present on the shortest paths were extracted and analyzed. Several of these genes have been reported to be related to lung cancer. However, some of them were not previously reported. Therefore, there are probably novel lung-cancer-related genes and have the potential to be biomarkers for diagnosis of lung cancer.

## 2. Materials and Methods

**2.1. Lung-Cancer-Related Gene List.** We compiled all 54 genes existing in the human nonsmall cell lung cancer (NSCLC) pathway and 84 genes in the small cell lung cancer (SCLC) pathway from KEGG [12]. These two gene sets and corresponding Ensembl protein IDs are listed in Additional file S1 in Supplementary Material available online at doi: <http://dx.doi.org/10.1155/2013/267375>.

**2.2. Lung Cancer Gene Expression Data.** The gene expression profiling in Kastner et al.’s work was used in our study [13], which includes 8 SCLC, 16 NSCLC, and 14 normal lung tissue samples. It was retrieved from NCBI Gene Expression Omnibus (GEO) (Accession number: GSE40275). The gene expression profile was obtained by the Human Exon 1.0 ST Array with 56283 probes corresponding to 26410 genes. Signal intensity was first log<sub>2</sub> transformed and then quantile normalized with “preprocessCore” package of R [14].

**2.3. Identifying Differentially Expressed Genes.** The “samr” package of R [15] was utilized to identify the differentially expressed genes between NSCLC, SCLC, and normal tissues separately with the criterion that false-discovery-rate- (FDR-) adjusted *P* value [16] was less than 0.01 and fold change was greater than 3 or less than 0.33.

**2.4. Cancer-Related Gene List.** A list of 742 cancer-related genes was compiled from three different sources [17]. First, a list of 457 cancer-related genes was collected from the Cancer Gene Census. Second, a list of cancer-related genes from the Atlas of Genetics and Cytogenetic in Oncology was retrieved [18]. We compiled the third list from the Human Protein Reference Database (HPRD) [19].

**2.5. STRING PPI Data and Shortest Path Identification.** The initial weighted PPI network was constructed based on data from STRING (version 9.0) [4] (<http://string.embl.de/>). Each interaction in STRING is evaluated by an interaction confidence score in range from 1 to 999 to quantify the likelihood that an interaction may occur. We used Dijkstra’s algorithm which has also been used in our previous works [20, 21] to identify the shortest path between each protein pair corresponding to the 54 NSCLC and 84 SCLC genes in the PPI network, respectively. Finally, all the proteins present on the shortest paths were ranked according to their betweenness. The Dijkstra’s algorithm was implemented with R package “igraph” [22].

**2.6. KEGG Pathway Enrichment Analysis.** KEGG pathway enrichment analysis was performed with the functional annotation tool DAVID [23]. The enrichment *P* value was corrected with the Benjamin multiple testing correction method to control family-wide false discovery rate less than 0.05 [24]. All the protein-coding genes in human genome were taken as background during the enrichment analysis.

## 3. Results

**3.1. Differentially Expressed Genes of the Gene Expression Profile.** With the SAMR method, 1918 significantly upexpressed probes and 2243 downexpressed probes corresponding to 1825 genes were identified for NSCLC when compared with 14 normal lung tissues (for probes see additional file S2, and for gene symbols see additional file S3). For SCLC, 819 significantly up-expressed probes and 820 down-expressed probes corresponding to 1063 genes were identified (for probes see additional file S2, and for gene symbols see additional file S3).

**3.2. Shortest Path Genes and Enrichment Analysis.** An undirected graph was constructed with the PPI data from STRING. Subsequently, we repeatedly chose a pair of proteins corresponding to 54 NSCLC genes and the 84 SCLC genes respectively, and the shortest path between these two proteins was determined with Dijkstra’s algorithm. A total of 1711 and 3916 shortest paths were obtained (see additional file S4) with lowest cost for NSCLC and SCLC containing 114 and 161 path genes, respectively. Shown in Figure 1 are the 1711 shortest paths between the 54 NSCLC genes. The weight was labeled on the edge between each of the interaction gene pairs. Shown in Figure 2 are the 3916 shortest paths between the 84 SCLC genes. To determine whether our 114 and 161 shortest path genes were also hubs in the background network, we performed a permutation to count the number

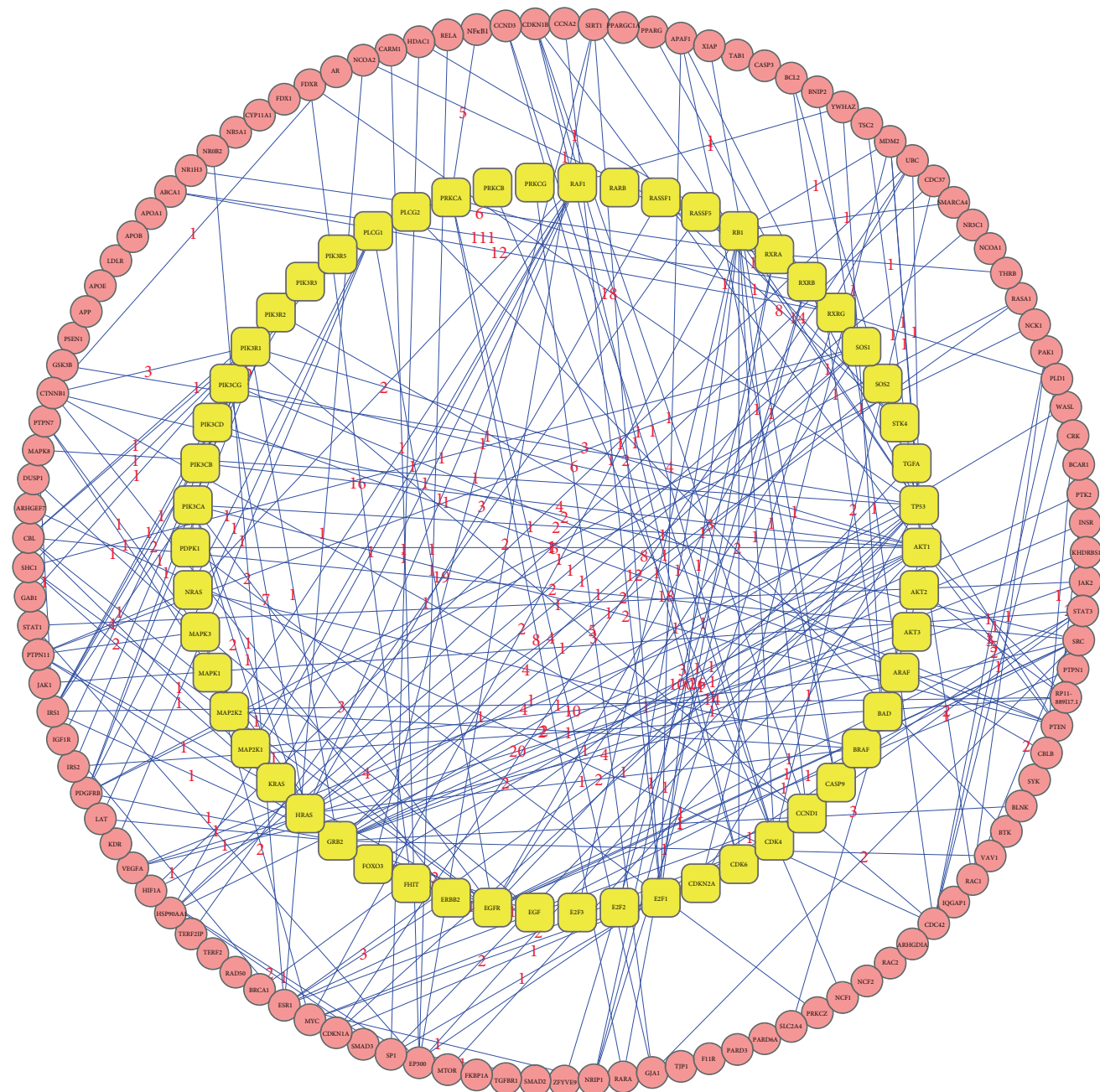


FIGURE 1: 1711 shortest paths between 54 NSCLC genes. The 1171 shortest paths between 54 NSCLC genes were identified with Dijkstra's algorithm based on PPI data from STRING. Yellow round represents 54 NSCLC genes. Red round represents 114 genes existing on shortest paths. Numbers on edges represent the edge weight to quantify the interaction confidence. The smaller the number, the stronger the interaction between two nodes.

of their occurrences on the shortest paths between 54 and 84 randomly selected genes only if they had a greater betweenness than that in our study. This process was repeated 2000 times, and the proportion of occurrences for the 114 and 161 shortest path genes was regarded as the *P* value. The detailed results thus obtained are given in additional file S5. Then we chose the 25 NSCLC and 38 SCLC shortest path genes with a *P* value less than 0.05 for further analysis (see additional file 5).

The GO enrichment analysis of 25 NSCLC shortest path genes indicated that they were significantly enriched in the regulation of intracellular signaling cascades and regulation of macromolecule metabolic processes (see additional file S6). These terms had been demonstrated to make great contributions to the survival and reproduction of cancer cells, while they also appeared in the enriched GO terms of 38 SCLC shortest path genes (see additional file S6). Besides these terms, the analysis result of SCLC shortest path



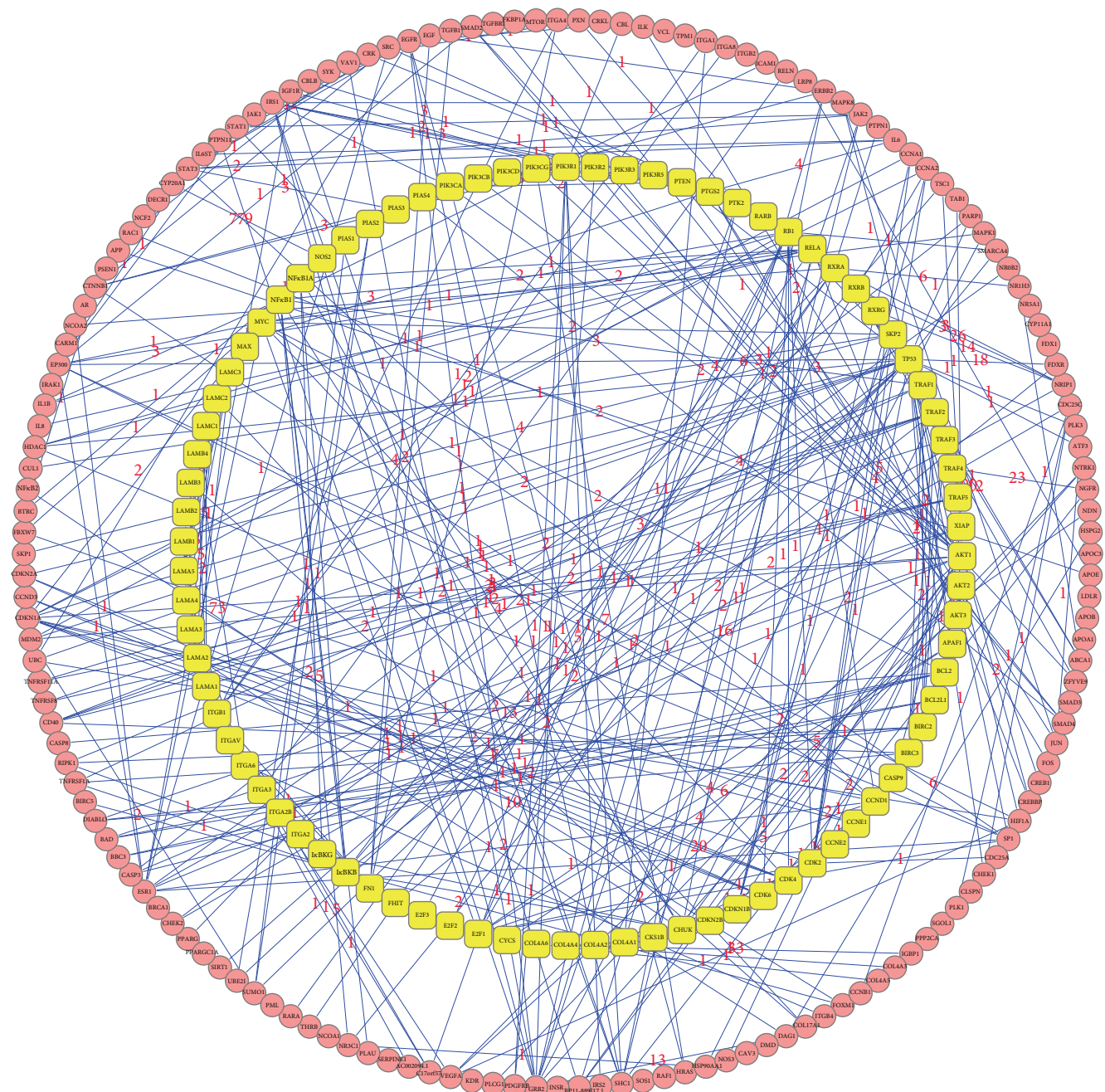


FIGURE 2: 3916 shortest paths between 84 SCLC genes. The 3916 shortest paths between 84 SCLC genes were identified with Dijkstra's algorithm based on PPI data from STRING. Yellow round represents 84 SCLC genes. Red round represents 161 genes existing on shortest paths. Numbers on edges represent the edge weight to quantify the interaction confidence. The smaller the number, the stronger the interaction between two nodes.

genes showed that they were significantly enriched in cell adhesion processes, suggesting that genes in this term might play an important role in differentiating SCLC from NSCLC (see additional file S6). The KEGG pathway enrichment of these 38 SCLC shortest path genes indicated that they were enriched in canonical-cancer-related pathways such as the cell cycle and p53 signaling pathway (Table 1).

**3.3. Comparing the Overlap between Candidate Genes with 742 Cancer-Related Genes.** The 25 and 38 shortest path genes were regarded as candidate genes for NSCLC and SCLC, respectively. We checked the overlap between 742 cancer genes and differentially expressed genes from the gene expression array as well as the overlap between the candidate genes identified in our study (Table 2). The entire 5-gene

TABLE 1: KEGG enrichment analysis of 38 SCLC shortest path genes.

Term	Count <sup>a</sup>	Percentage <sup>b</sup>	<i>P</i> value	Benjamini adjusted <i>P</i> value
Focal adhesion	8	21.1	$1.40E - 05$	$6.70E - 04$
Regulation of actin cytoskeleton	7	18.4	$2.20E - 04$	$5.40E - 03$
Arrhythmogenic right ventricular cardiomyopathy (ARVC)	5	13.2	$2.70E - 04$	$4.40E - 03$
ECM-receptor interaction	5	13.2	$4.00E - 04$	$4.80E - 03$
Hypertrophic cardiomyopathy (HCM)	5	13.2	$4.20E - 04$	$4.10E - 03$
Dilated cardiomyopathy	5	13.2	$5.70E - 04$	$4.60E - 03$
Cell cycle	5	13.2	$1.80E - 03$	$1.20E - 02$
p53-signaling pathway	4	10.5	$2.90E - 03$	$1.70E - 02$

<sup>a</sup>The number of genes belonging to a certain pathway.

<sup>b</sup>The percentage of genes belonging to a certain pathway accounts for all the genes undergoing KEGG pathway analysis.

TABLE 2: Overlap between candidate genes and cancer-related genes.

Gene set	Number of candidate genes	Overlap with 742 cancer genes	<i>P</i> value
NSCLC from array	1825	93	$6.698e - 04$
SCLC from array	1063	69	$2.218e - 06$
NSCLC in our study	25	6	$2.518e - 05$
SCLC in our study	38	5	$2.559e - 03$

*P* value was calculated with the hypergeometric test assuming the total number of protein-coding genes was 20000.

set can be found in additional file S3. From Table 2, we can see that both the lung cancer candidate genes identified from the gene expression array and those identified by our method had a significant overlap with the 742 cancer genes. However, the 25 NSCLC candidate genes identified with our method contained more cancer genes than those from the gene expression array ( $P$  value =  $3.858e - 03$ ) (Table 3). The 38 SCLC candidate genes had a higher percentage of cancer-related genes (0.1316) than those from expression array (0.0649) though the  $P$  value of Fisher's exact test was not significant ( $P$  value = 0.186). At least, the 38 SCLC candidate genes contained comparable cancer-related genes as those from gene expression array.

## 4. Discussion

**4.1. Shortest Path Genes in Non-small Cell Lung Cancer (NSCLC).** We identified 25 shortest path genes in NSCLC and 38 shortest path genes in SCLC with a permutation  $P$  value less than 0.05. Intriguingly the top five shortest path genes in NSCLC are also among the most significant genes in SCLC, while SCLC has several unique genes with large betweenness values. These may help to reveal the relationship between the two major types of lung cancer.

As in NSCLC, HSP90AA1 [25–27] has been well documented to be relevant to lung cancer. We focus on candidate genes with large betweenness values and discuss the potential relationship between them and lung cancer.

TABLE 3: Comparing the overlap between candidate genes with cancer-related genes.

Gene set	Number of candidate genes	Overlap with 742 cancer genes	<i>P</i> value
NSCLC from array	1825	93	
NSCLC in our study	25	6	$3.858e - 03$
SCLC from array	1063	69	
SCLC in our study	38	5	0.186

*P* value was calculated with Fisher's exact test.

Estrogen receptor 1 (ESR1) belongs to the nuclear steroid hormone receptor superfamily which acts as ligand-dependent, sequence-specific transcription factors and regulates the expression of genes involved in signal transduction, cell-cycle control, and cell survival [28]. Previous evidence showed that the proportion of never smokers among women with lung cancer is higher compared with men. Hypermethylation of ESR1 was reported to be detected only in lung tumors, but not in normal lung tissues, with a higher frequency being found in male patients than in female patients [29]. These all indicated ESR1 as a prognostic factor in lung cancer and as a potential target of hormone therapy.

ATP-binding cassette sub-family A member 1 (ABCA1) is a sulfonylurea-sensitive and cAMP-dependent anion transporter with critical impact on intracellular cholesterol transport. Cholesterol level increase has been found in cancers compared with normal tissue in many kinds of cancers [30], such as oral cancer [31]. Smith and Land demonstrated in colon cancer cells that ABCA1 had an anticancer activity in which deficiency allowed for increased mitochondrial cholesterol, inhibited release of mitochondrial cell death-promoting molecules, and facilitated cancer cell survival [32]. As abnormal metabolism is generally found in cancer, ABCA1 deserves further investigation with regard to its role in lung cancer.

Insulin receptor substrate 1 (IRS1) is an adaptor protein for insulin-like growth factor (IGF) signaling and is associated with IGF-stimulated proliferation [33]. It has been reported to be downregulated in NSCLC [34], and

its degradation accelerates lung tumor growth by upgrading interaction between the potent mitogen platelet-derived growth factor receptor (PDGFR) and phosphatidylinositol 3 kinase (PI3 K) [35]. Correspondingly, our study shows that the shortest path of IRS1 is designated more than 100 and is significant in both NSCLC genes and SCLC genes, indicating that it may play a crucial part in lung cancer development.

FDXR (NADPH: adrenodoxin oxidoreductase) serves as the first electron transfer protein in the mitochondrial P450 systems. FDXR is identified to be target of the p53 family. It could be induced in a p53-dependent way by DNA damage in cells and participated in p53-mediated apoptosis via generating oxidative stress in mitochondria [36, 37]. Owing to the significance of p53 in apoptosis during tumorigenesis, the contribution of FDXR to lung cancer is worthy of further elucidation.

**4.2. Shortest Path Genes in Small Cell Lung Cancer (SCLC).** The KEGG pathway enrichment analysis shows that there is a distinct group of shortest path genes in SCLC compared with NSCLC. These are the extracellular-matrix- (ECM-) related genes (Table 1). This coincides with the KEGG pathway enrichment analysis result of known SCLC pathway genes. ECM surrounds SCLC cells and includes collagen IV, tenascin, fibronectin, and laminin. Cell surface receptor integrins interact with ECM components and numerous signal transduction pathways which play important roles in cell cycle regulation, apoptosis, and so on and thus promote cancer cell proliferation [38]. Hodgkinson et al. found that ECM can inhibit the caspase-3 activation and subsequent cell apoptosis induced by etoposide via stimulating phosphatidylinositol 3-kinase- (PI3K-) signaling pathway in SCLC cells in a ITGB1/PI3K-dependent way [39]. Choi et al. demonstrated that downregulation of the phosphorylation activity of ILK (integrin-linked kinase) by single deletion of ILK protein itself or deletion of ITGB4/ILK complex could suppress the invasion of ovarian cancer [40]. Other studies also demonstrate that the intracellular signals activated by ECM components account for the high metastasis potential and drug resistance of SCLC [41]. In this work, we found that collagen IV members COL4A5 and COL4A3, integrin members ITGA1, ITGB4 and ITGA4, and linked kinase ILK all have a betweenness of more than 80 and a  $P$  value  $< 0.05$ , all of which may indicate their crucial roles in SCLC.

Forkhead box protein M1 (FOXO1) is a transcription factor regulating cell proliferation and DNA damage repair [42, 43]. Research shows that it could be phosphorylated by MAPK (ERK) kinase [44] and then activate the expression of a number of cell-cycle-related genes which are crucial for DNA replication and mitotic division in the Ras-mitogen-activated protein-kinase- (MAPK-) signaling pathway, such as cyclin A2, cyclin B1, Aurora B kinase, Cdc25B phosphatase- and Polo-like kinase1 [45]. Additionally, the protein level of FOXO1 has been found increased in prostate adenocarcinomas [46], infiltrating ductal breast carcinomas [47], basal cell carcinomas [48], intrahepatic cholangiocarcinomas [49], and in many other solid tumors [50]. A study by Kim et al. [51] showed that in human NSCLC Foxo1 protein is over-expressed and promotes tumor cells proliferation during the

TABLE 4: The functional similarity between identified lung cancer genes and 742 cancer genes.

	742 cancer genes
1825 NSCLC genes from array	0.4314*
1063 SCLC genes from array	0.4845*
25 NSCLC genes from our study	0.5554*
38 SCLC genes from our study	0.6919*

\* Pearson correlation coefficient of functional profiles.

development of NSCLC. These all indicate that FOXO1 may play an import role in SCLC as well.

Immunoglobulin-binding protein 1 (IGBP1) was formerly identified as a signal transduction molecule with a surface IgM receptor. More recently, it has been shown to regulate the phosphatase catalytic activity of protein phosphatase 2A (PP2A) [52]. PP2A is composed of a majority of cellular serine/threonine phosphatases [53] and regulates a number of important cellular processes, such as cell cycle transition, apoptosis, transcription, translation, autophagy [54], and cell transformation [55]. IGBP1 directly interacts with the catalytic subunit of PP2A [56], and this interaction leads to an antiapoptosis function. Recent studies show that, in carcinogen-transformed human cells and primary human cancers such as primary lung cancers, primary hepatocellular carcinomas and primary breast cancers, the expression level of IGBP1 is upregulated, [57]. Sakashita et al. found its overexpression in small cell adenocarcinomas [58], and Li et al. found that in a lung adenocarcinoma cell line the interaction of IGBP1 and Lactoferrin could induce cell apoptosis [59], implying IGBP1 to be a candidate target for SCLC therapy.

**4.3. Functional Similarities between Candidate Genes and Known Cancer Genes.** In order to compare the functional similarities between our candidate genes and the 742 known cancer genes, their functional profiles were constructed using the  $-\log_{10}$  of the hypergeometric test  $P$  value on Gene Ontology (GO) terms [20, 21]. Then the Pearson correlation coefficient of their functional profiles was calculated [20, 21]. The functional similarities of five gene sets are shown in Table 4. All five gene sets can be found in additional file S3. Our 25 NSCLC (0.5554) and 38 SCLC (0.6919) candidate genes both had greater functional similarity with the cancer genes than the NSCLC (0.43139) and SCLC (0.48451) genes identified from gene expression profiles. It is suggested that our way is more efficient in identifying cancer-related genes.

## 5. Conclusion

In this study, we propose a computational method based on a protein-protein interaction network to identify cancer-related genes. We applied this method to lung cancer to find the shortest paths between 54 NSCLC and 84 SCLC genes in the protein-protein interaction network constructed based on STRING data and selected the 25 and 38 genes with a significant  $P$  value for NSCLC and SCLC, respectively.



Analysis of these shortest path genes indicates that some of these genes, such as ESRI, FDXR, ABCA1, IRS1, HSP90AA1, FOXM1, and IGBP1 are related to lung cancer. In addition, the candidate genes of lung cancer identified in our study contain more cancer genes than those identified from gene expression profiles. Moreover, it is revealed that our candidate genes have greater functional similarity with the cancer genes than those identified from gene expression profiles. These candidate genes may be worth experiment validation and further research. It is expected that this method is useful in predicting novel cancer-related genes and has widespread use in cancer research.

## Authors' Contribution

B.-Q. Li and J. You contributed equally to this work.

## Acknowledgments

This work was supported by grants from National Basic Research Program of China (2011CB510102 and 2011CB510101), Innovation Program of Shanghai Municipal Education Commission (12ZZ087), and the Grant of "The First-class Discipline of Universities in Shanghai."

## References

- [1] R. Siegel, D. Naishadham, and A. Jemal, "Cancer statistics," *A Cancer Journal for Clinicians*, vol. 62, pp. 10–29, 2012.
- [2] J. P. van Meerbeeck, D. A. Fennell, and De Ruysscher, "DKM Small-cell lung cancer," *The Lancet*, vol. 378, pp. 1741–1755.
- [3] E. Nabieva, K. Jim, A. Agarwal, B. Chazelle, and M. Singh, "Whole-proteome prediction of protein function via graph-theoretic analysis of interaction maps," *Bioinformatics*, vol. 21, supplement 1, pp. i302–i310, 2005.
- [4] D. Szklarczyk, A. Franceschini, M. Kuhn et al., "The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored," *Nucleic Acids Research*, vol. 39, no. 1, pp. D561–D568, 2011.
- [5] R. Sharan, I. Ulitsky, and R. Shamir, "Network-based prediction of protein function," *Molecular Systems Biology*, vol. 3, article 88, 2007.
- [6] P. Bogdanov and A. K. Singh, "Molecular function prediction using neighborhood features," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 7, no. 2, pp. 208–217, 2010.
- [7] Y. A. Kourmpetis, A. D. van Dijk, M. C. Bink, R. C. van Ham, and C. J. ter Braak, "Bayesian Markov Random Field analysis for protein function prediction based on network data," *PLoS One*, vol. 5, article e9293, 2010.
- [8] K. L. Ng, J. S. Ciou, and C. H. Huang, "Prediction of protein functions based on function-function correlation relations," *Computers in Biology and Medicine*, vol. 40, no. 3, pp. 300–305, 2010.
- [9] U. Karaöz, T. M. Murali, S. Letovsky et al., "Whole-genome annotation by using evidence integration in functional-linkage networks," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 101, no. 9, pp. 2888–2893, 2004.
- [10] S. Letovsky and S. Kasif, "Predicting protein function from protein/protein interaction data: a probabilistic approach," *Bioinformatics*, vol. 19, supplement 1, pp. i197–i204, 2003.
- [11] E. W. Dijkstra, "A note on two problems in connexion with graphs," *Numerische Mathematik*, vol. 1, no. 1, pp. 269–271, 1959.
- [12] M. Kanehisa and Subramaniam, "The KEGG database," *Novartis Foundation Symposium*, vol. 247, pp. 91–103, 2002.
- [13] S. Kastner, T. Voss, S. Keuerleber, C. Glickel, M. Freissmuth, and W. Sommergruber, "Expression of G protein-coupled receptor 19 in human lung cancer cells is triggered by entry into S-phase and supports G(2)-m cell-cycle progression," *Molecular Cancer Research*, vol. 10, pp. 1343–1358, 2012.
- [14] B. M. Bolstad, R. A. Irizarry, M. Åstrand, and T. P. Speed, "A comparison of normalization methods for high density oligonucleotide array data based on variance and bias," *Bioinformatics*, vol. 19, no. 2, pp. 185–193, 2003.
- [15] S. Zhang, "A comprehensive evaluation of SAM, the SAM R-package and a simple modification to improve its performance," *BMC Bioinformatics*, vol. 8, article 230, 2007.
- [16] Y. Benjamini and Y. Hochberg, "Controlling the false discovery rate: a practical and powerful approach to multiple testing," *Journal of the Royal Statistical Society Series B*, vol. 57, pp. 289–300, 1995.
- [17] S. H. Nagaraj and A. Reverter, "A Boolean-based systems biology approach to predict novel genes associated with cancer: application to colorectal cancer," *BMC Systems Biology*, vol. 5, article 35, 2011.
- [18] J. L. Huret, P. Dessen, and A. Bernheim, "Atlas of genetics and cytogenetics in oncology and haematology, year 2003," *Nucleic Acids Research*, vol. 31, no. 1, pp. 272–274, 2003.
- [19] T. S. K. Prasad, R. Goel, K. Kandasamy et al., "Human protein reference database—2009 update," *Nucleic Acids Research*, vol. 37, no. 1, pp. D767–D772, 2009.
- [20] B. Q. Li, J. Zhang, T. Huang, L. Zhang, and Y. D. Cai, "Identification of retinoblastoma related genes with shortest path in a protein-protein interaction network," *Biochimie*, vol. 94, pp. 1910–1917, 2012.
- [21] B. Q. Li, T. Huang, L. Liu, Y. D. Cai, and K. C. Chou, "Identification of colorectal cancer related genes with mRMR and shortest path in protein-protein interaction network," *PLoS One*, vol. 7, Article ID e33393, 2012.
- [22] G. Csardi and T. Nepusz, "The igraph software package for complex network research," *InterJournal Complex Systems*, 2006.
- [23] D. W. Huang, B. T. Sherman, and R. A. Lempicki, "Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources," *Nature Protocols*, vol. 4, no. 1, pp. 44–57, 2009.
- [24] Y. Benjamini and D. Yekutieli, "The control of the false discovery rate in multiple testing under dependency," *Annals of Statistics*, vol. 29, no. 4, pp. 1165–1188, 2001.
- [25] L. Whitesell and S. L. Lindquist, "HSP90 and the chaperoning of cancer," *Nature Reviews Cancer*, vol. 5, no. 10, pp. 761–772, 2005.
- [26] J. Trepel, M. Mollapour, G. Giaccone, and L. Neckers, "Targeting the dynamic HSP90 complex in cancer," *Nature Reviews Cancer*, vol. 10, no. 8, pp. 537–549, 2010.
- [27] M. Senju, N. Sueoka, A. Sato et al., "Hsp90 inhibitors cause G2/M arrest associated with the reduction of Cdc25C and Cdc2 in lung cancer cell lines," *Journal of Cancer Research and Clinical Oncology*, vol. 132, no. 3, pp. 150–158, 2006.
- [28] J. Frasor, J. M. Danes, B. Komm, K. C. N. Chang, C. Richard Lytle, and B. S. Katzenellenbogen, "Profiling of estrogen up- and

- down-regulated gene expression in human breast cancer cells: insights into gene networks and pathways underlying estrogenic control of proliferation and cell phenotype," *Endocrinology*, vol. 144, no. 10, pp. 4562–4574, 2003.
- [29] J. C. Lai, Y. W. Cheng, H. L. Chiou, M. F. Wu, C. Y. Chen, and H. Lee, "Gender difference in estrogen receptor alpha promoter hypermethylation and its prognostic value in non-small cell lung cancer," *International Journal of Cancer*, vol. 117, no. 6, pp. 974–980, 2005.
- [30] Y. Yoshioka, J. Sasaki, M. Yamamoto et al., "Quantitation by <sup>1</sup>H-NMR of dolichol, cholesterol and choline-containing lipids in extracts of normal and pathological thyroid tissue," *NMR in Biomedicine*, vol. 13, pp. 377–383, 2000.
- [31] K. Kolanjiappan, C. R. Ramachandran, and S. Manoharan, "Biochemical changes in tumor tissues of oral cancer patients," *Clinical Biochemistry*, vol. 36, no. 1, pp. 61–65, 2003.
- [32] B. Smith and H. Land, "Anticancer activity of the cholesterol exporter ABCA1 gene," *Cell Reports*, vol. 2, pp. 580–590.
- [33] D. Yee, "Targeting insulin-like growth factor pathways," *British Journal of Cancer*, vol. 94, no. 4, pp. 465–468, 2006.
- [34] C. H. Han, J. Y. Cho, J. T. Moon et al., "Clinical significance of insulin receptor substrate-1 down-regulation in non-small cell lung cancer," *Oncology Reports*, vol. 16, no. 6, pp. 1205–1210, 2006.
- [35] A. M. Houghton, D. M. Rzymkiewicz, H. Ji et al., "Neutrophil elastase-mediated degradation of IRS-1 accelerates lung tumor growth," *Nature Medicine*, vol. 16, no. 2, pp. 219–223, 2010.
- [36] G. Liu and X. Chen, "The ferredoxin reductase gene is regulated by the p53 family and sensitizes cells to oxidative stress-induced apoptosis," *Oncogene*, vol. 21, no. 47, pp. 7195–7204, 2002.
- [37] P. M. Hwang, F. Bunz, J. Yu et al., "Ferredoxin reductase affects p53-dependent, 5-fluorouracil-induced apoptosis in colorectal cancer cells," *Nature Medicine*, vol. 7, no. 11, pp. 1111–1117, 2001.
- [38] A. J. Ridley, M. A. Schwartz, K. Burridge et al., "Cell migration: integrating signals from front to back," *Science*, vol. 302, no. 5651, pp. 1704–1709, 2003.
- [39] P. S. Hodgkinson, T. Elliott, W. S. Wong et al., "ECM overrides DNA damage-induced cell cycle arrest and apoptosis in small-cell lung cancer cells through  $\beta$ 1 integrin-dependent activation of PI3-kinase," *Cell Death and Differentiation*, vol. 13, no. 10, pp. 1776–1788, 2006.
- [40] Y. P. Choi, B. G. Kim, M. Q. Gao, S. Kang, and N. H. Cho, "Targeting ILK and  $\beta$ 4 integrin abrogates the invasive potential of ovarian cancer," *Biochemical and Biophysical Research Communications*, vol. 427, no. 3, pp. 642–648, 2012.
- [41] T. Sethi, R. C. Rintoul, S. M. Moore et al., "Extracellular matrix proteins protect small cell lung cancer cells against apoptosis: a mechanism for small cell lung cancer growth and drug resistance in vivo," *Nature Medicine*, vol. 5, no. 6, pp. 662–668, 1999.
- [42] Y. Tan, P. Raychaudhuri, and R. H. Costa, "Chk2 mediates stabilization of the FoxM1 transcription factor to stimulate expression of DNA repair genes," *Molecular and Cellular Biology*, vol. 27, no. 3, pp. 1007–1016, 2007.
- [43] Z. Fu, L. Malureanu, J. Huang et al., "Plk1-dependent phosphorylation of FoxM1 regulates a transcriptional programme required for mitotic progression," *Nature Cell Biology*, vol. 10, no. 9, pp. 1076–1082, 2008.
- [44] R. Y. M. Ma, T. H. K. Tong, A. M. S. Cheung, A. C. C. Tsang, W. Y. Leung, and K. M. Yao, "Raf/MEK/MAPK signaling stimulates the nuclear translocation and transactivating activity of FOXM1c," *Journal of Cell Science*, vol. 118, no. 4, pp. 795–806, 2005.
- [45] R. H. Costa, V. V. Kalinichenko, M. L. Major, and P. Raychaudhuri, "New and unexpected: forkhead meets ARF," *Current Opinion in Genetics and Development*, vol. 15, no. 1, pp. 42–48, 2005.
- [46] T. V. Kalin, I. C. Wang, T. J. Ackerson et al., "Increased levels of the FoxM1 transcription factor accelerate development and progression of prostate carcinomas in both TRAMP and LADY transgenic mice," *Cancer Research*, vol. 66, no. 3, pp. 1712–1720, 2006.
- [47] D. R. Wonsey and M. T. Follettie, "Loss of the forkhead transcription factor FoxM1 causes centrosome amplification and mitotic catastrophe," *Cancer Research*, vol. 65, no. 12, pp. 5181–5189, 2005.
- [48] M. T. Teh, S. T. Wong, G. W. Neill, L. R. Ghali, M. P. Philpott, and A. G. Quinn, "FOXO1 is a downstream target of Gli1 in basal cell carcinomas," *Cancer Research*, vol. 62, no. 16, pp. 4773–4780, 2002.
- [49] K. Obama, K. Ura, M. Li et al., "Genome-wide analysis of gene expression in human intrahepatic cholangiocarcinoma," *Hepatology*, vol. 41, no. 6, pp. 1339–1348, 2005.
- [50] C. Pilarsky, M. Wenzig, T. Specht, H. D. Saeger, and R. Grützmann, "Identification and validation of commonly over-expressed genes in solid tumors by comparison of microarray data," *Neoplasia*, vol. 6, no. 6, pp. 744–750, 2004.
- [51] I. M. Kim, T. Ackerson, S. Ramakrishna et al., "The Forkhead Box m1 transcription factor stimulates the proliferation of tumor cells during development of lung cancer," *Cancer Research*, vol. 66, no. 4, pp. 2153–2161, 2006.
- [52] M. Kong, D. Ditsworth, T. Lindsten, and C. B. Thompson, " $\alpha$ 4 is an essential regulator of PP2A phosphatase activity," *Molecular Cell*, vol. 36, no. 1, pp. 51–60, 2009.
- [53] V. Janssens and J. Goris, "Protein phosphatase 2A: a highly regulated family of serine/threonine phosphatases implicated in cell growth and signalling," *Biochemical Journal*, vol. 353, no. 3, pp. 417–439, 2001.
- [54] T. Yorimitsu, C. He, K. Wang, and D. J. Klionsky, "Tap42-associated protein phosphatase type 2A negatively regulates induction of autophagy," *Autophagy*, vol. 5, no. 5, pp. 616–624, 2009.
- [55] J. D. Arroyo and W. C. Hahn, "Involvement of PP2A in viral and cellular transformation," *Oncogene*, vol. 24, no. 52, pp. 7746–7755, 2005.
- [56] K. Murata, J. Wu, and D. L. Brautigan, "B cell receptor-associated protein  $\alpha$ 4 displays rapamycin-sensitive binding directly to the catalytic subunit of protein phosphatase 2A," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 94, no. 20, pp. 10624–10629, 1997.
- [57] L. P. Chen, Y. D. Lai, D. C. Li et al., " $\alpha$ 4 is highly expressed in carcinogen-transformed human cells and primary human cancers," *Oncogene*, vol. 30, no. 26, pp. 2943–2953, 2011.
- [58] S. Sakashita, D. Li, N. Nashima et al., "Overexpression of immunoglobulin (CD79a) binding protein1 (IGBP-1) in small lung adenocarcinomas and its clinicopathological significance," *Pathology International*, vol. 61, no. 3, pp. 130–137, 2011.
- [59] D. Li, S. Sakashita, Y. Morishita et al., "Binding of lactoferrin to IGBP1 triggers apoptosis in a lung adenocarcinoma cell line," *Anticancer Research*, vol. 31, no. 2, pp. 529–534, 2011.



