Integrated bioinformatics analysis reveals ASPM and CENPF with prognostic value in lung cancer

jinghang li

the first affiliated hospital of Nanjing medical university

Jing Zhang

the first affiliated hospital of Nanjing medical university

Lin Huang

the first affiliated hospital of Nanjing medical university

Sheng Zhao (✉ zhaosheng0824@njmu.edu.cn)

<https://orcid.org/0000-0002-4602-1806>

Research article

Keywords: Lung cancer; bioinformatical analysis; candidate gene identification; genetic alteration; network analysis

Posted Date: October 15th, 2019

DOI: <https://doi.org/10.21203/rs.2.13321/v2>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.
[Read Full License](#)

Abstract

Lung cancer (LC) is the most frequent type of cancer in the world. But the mechanism of LC is still largely unknown. In this study, we analyzed three lung cancer gene expression microarrays of different pathologic types to explore the potential candidate genes in LC by Integrated bioinformatical methods. 459 overlapped differentially expressed genes (DEGs) were explored in three GEO gene expression profiles of different pathologic types of lung cancer and function annotation of DEGs were performed. The main biological process of DEGs was regulation of vasculature development and angiogenesis. The most significant molecular function of DEGs was TGF- β receptor activity. The most significant Reactome pathway of DEGs was cell cycle and extracellular matrix organization pathway. The PPI network of the DEGs was constructed and 23 candidate hub genes were identified in the network. Kaplan-Meier survival analysis show 21 genes were associated with the prognosis of LC. The genetic alterations analysis of these genes by using cBioPortal shown ASPM has the highest genetic alteration rate of 9% in main pathological types of 3191 LC patients, CENPF has the second highest alteration rate of 6% in LC patients. ASPM and CENPF also identified have a significant co-occurrence relationship in LC, and the GO analysis shown they both participate in the regulation of cell cycle. In the TF-miRNA-gene network of 21 genes shown CENPF have the most significant value in the network and the most relevant TF are NFYA, E2F1 and MYC. In conclusion, this study explored several key genes about LC and analyzed potential TF of those genes, provides possible therapeutic targets and biomarker for further clinical application.

Background

Lung cancer (LC), one of the highest incidence malignant tumors worldwide, is one of the leading cause of cancer-associated death in the world(1). The general mortality of LC was increasing rapidly in recent years, from 3.5 million in 1990 to 4.2 million in 2015(2). Due to the difficulties in early diagnosis for LC and lacking of effective medical therapy, the improvement of long-term survival rate of LC patients is still too slow(1). The occurrence and progression of lung cancer, likes most types of human tumors, is identified as a heterogeneous progress, numerous factors such as air pollution, smoking and genetic factors are all important causes of lung cancer(3, 4). The abnormal expression and alteration of genes is vital for the development of LC(5, 6), and therapy methods aimed at genetic alteration improved the survival of terminal lung cancer patients in recent years(7, 8). Thus, it is very important to investigate the key genes and explore the precise mechanisms of lung cancer, and this may be helpful to find biomarkers for early diagnosis of LC and find novel therapeutic targets for lung cancer.

The rapid progresses in gene-expression microarrays and bioinformatics have promoted the discovery of key genes and underlying molecular mechanisms in human diseases especially in tumors(9-11). In recent years, numerous microarray studies of LC have been carried out and produced a large amount of chip data mainly instored in public database like Gene Expression Omnibus database (GEO) and The Cancer Genome Atlas (TCGA). Most microarray studies of LC are single cohort study and have a small sample size, and not performed according to the pathological types of lung tumor. All these effects cause a poor reliability of the results. However, using integrated bioinformatics methods to integrate different studies

increased the sample size and analyzing microarray data according to the pathological types of lung tumor might reduce these disadvantages.

In this study, we selected three original human lung cancer microarray datasets from GEO, GSE74706 including 10 LAC lung tissues, 8 LSCC lung tissues and 18 normal lung tissues, GSE60052 including 79 SCLC lung tissues and 7 normal control lung tissues, GSE19804 including 60 normal control lung tissues and 60 lung cancer tissues. Then we analyzed DEGs by using online tools GEOR2 and BioJupies (12) according to the pathological type of lung tumor, and identified overlapped DEGs in all three pathological type, and then performed GO term and pathway enrichment analysis, and explored key candidate genes in the PPI network by using Cytoscape software. We further performed Kaplan-Meier survival analysis of the hub genes use online tool Kaplan-Meier Plotter (<http://kmplot.com/analysis/>). The cBioPortal for Cancer Genomics (<http://www.cbioportal.org/>) (13) was used to explore the genetic alterations and co-occurrence relationship of the hub genes in LC patients. In the end, a TF-miRNA-gene network was constructed by NetworkAnalyst ([https:// www. networkanalyst.ca/](https://www.networkanalyst.ca/)) (14) to explore potential regulatory mechanism of these genes.

Results

1. Identification of DEGs in Different Pathologic Types of LC

We screened 1328 DEGs of lung cancer from GSE19804, 4189 DEGs of LAC and 5764 DEGs of LSCC from GSE74706, 3368 DEGs of SCLC from GSE60052, with cut-off criterion of adjust P value ≤ 0.05 and $|\log FC| \geq 1$. And 459 overlapped DEGs were obtained (Figure 1).

2. Functional Annotation and Pathways Enrichment of DEGs

Gene Ontology analysis and Reactome pathway analysis of DEGs were performed by Enrichr. In the Gene Ontology analysis, the DEGs were enriched into three groups, including biological process, molecular function and cellular component (Figure 2). In the biological process, the DEGs mainly enriched in positive regulation of vasculature development, regulation of angiogenesis and positive regulation of angiogenesis. In the molecular function, the DEGs mainly enriched in transforming growth factor beta-activated receptor activity, transmembrane receptor protein serine/threonine kinase activity, protein homodimerization activity and integrin binding. In the cellular component, the DEGs mainly enriched in condensed nuclear chromosome, centromeric region, amellar body, integral component of plasma membrane and condensed chromosome kinetochore. In the Reactome pathway analysis, the most significant pathway was Cell Cycle and Mitotic, Cell Cycle and Hemostasis, Platelet degranulation, Extracellular matrix organization and signaling by Rho GTPases.

3. PPI Network Construction and Key Candidate Genes Screening

All the 459 DEGs were filtered into the PPI network (Figure 3). We used two kinds of topological algorithms (Degree, MCC) to calculate the top 25 nodes (Figure 4A-B, Table 1A-B), and we selected 23

overlapped nodes (CDK1, CDC20, AURKA, UBE2C, CCNA2, KIF11, MKI67, BUB1B, TOP2A, PBK, MAD2L1, TPX2, BIRC5, ASPM, RRM2, BUB1, DLGAP5, TTK, KIF2C, NUSAP1, NCAPG, CENPF, MELK) as candidate genes for further study (Figure 4C).

4. Kaplan-Meier Survival Analysis of the Candidate Genes

We further performed Kaplan-Meier survival analysis of these candidate genes in 2437 lung cancer patients by using online tool Kaplan-Meier Plotter. And discovered 21 genes (ASPM, AURKA, BIRC5, BUB1, CCNA, CDC20, CENPF, CT84, CT96, DLGAP5, KIF2C, KIF11, MAD2L1, MEIK, MK167, NCAPG, RRM2, SSK1, TOP2A, TPX2, UBE2C) have significant association with the prognosis of lung cancer in this analysis (Figure.5).

5. The Genetic Alteration Analysis of Candidate Genes in LC patients

We examined the genetic alterations of these 21 genes by using cBioPortal. The results shown 19 genes have genetic alterations in the main pathological type of lung tumor (Figure 6), and amongst these genes ASPM has the highest genetic alteration rate of 9% in 3191 LC patients, and CENPF has the second highest alteration rate of 6%. ASPM and CENPF also have a significant co-occurrence relationship in LC (Table.2).

6. Transcription Factor (TF) -miRNA-Gene Network of the Candidate Genes

In the TF -miRNA-gene network of 21 genes, CENPF have the most significant value in the network, and the most relevant TF are NFYA, E2F1 and MYC (Figure 7). There was no significant miRNA associated with these genes in this network.

7. Function and Pathway Analysis of ASPM and CENPF

The PPI network of ASPM shown ASPM and CENPF had significant interaction (Figure 8), and the related function and pathway of these two genes were both enriched in cell cycle (Supplementary Tables.1-4).

Discussion

There was a great progress in targeted therapy and immunotherapy for lung cancers in recent years, but the overall mortality of lung cancer is still high. It is critical to early diagnosis and treatment for lung cancer. And it is important to seeking for novel therapy targets and biomarkers for the prevention, diagnose and treatment of lung cancer. Bioinformatic analysis and microarray expression profiling analysis is considered as a powerful, comprehensive and accurate method to discover novel diagnosis markers or therapeutic targets for various diseases, particularly for cancers.

In this study, we integrated three lung cancer profile datasets from GEO database, and analyzed the data according to the pathologic types of lung cancer. 459 overlapped DEGs have been identified in three main

pathologic types of lung cancer, including LAC, LSCC and SCLC. GO term enrichment analysis was performed for annotating DEGs and the result demonstrated that the significant biological process related to cancer was regulation of vasculature development and angiogenesis.

Angiogenesis is essential for tumor growth and metastasis, and play an important role in the control of cancer progression. Angiogenesis has been validated as an effective therapeutic target in several kinds of tumors including lung cancers through randomized controlled clinical trials, and one of these effective drug is vascular endothelial growth factor (VEGF) pathway inhibitors(15, 16).

The significant molecular function of the DEGs was TGF- β receptor activity. The TGF- β pathway is important in the genesis and development of tumor. In the initiation stage of cancer development, TGF- β acts as a potential tumor suppressor and play a critical role in inhibiting the genesis of tumor by suppressing the proliferation of tumor epithelial cells. As in the stage of tumor progresses, TGF- β transfer into a tumor promoter and participates in the process of cancer progress by promoting the proliferation, invasion and metastasis of cancer cells and also vital for keeping the potentiality of cancer stem cell(17, 18).

In the Reactome pathway analysis, the most significant pathway of DEGs was cell cycle and extracellular matrix organization pathway. Cell cycle is a fundamental process of cell life, which controls the growth and proliferation of cell. Abnormal cell cycle is strongly associated with cell carcinogenesis. Cyclin-dependent kinases(Cdks), a key factors that regulate the cell cycle, is crucial in cell cycle, and is an ideal therapy target in tumor(19). Extracellular matrix (ECM) is an important tissue barrier to prevent tumor metastasis. The key components of ECM, fibronectin and laminin, are connected to the surface membrane integrin receptor of cancer cell, and also determines the shape of cancer cell and controls the differentiation and migration of cancer cell(20).

The PPI network of DEGs were constructed and 23 key candidate genes were identified, and 21 genes have significant association with the prognosis of lung cancer. Amongst these genes, ASPM had the highest genetic alteration rate of 9% in 3191 LC patients of three main pathological types, and CENPF has the second highest alteration rate of 6%. Data shown that ASPM and CENPF had a significant co-occurrence relationship in LC and these two genes have an interaction in the PPI network, the related function and pathway of these two genes are both enriched in cell cycle. The main genetic alteration of ASPM and CENPF was amplification and missense mutation.

ASPM—abnormal spindle microtubule assembly—encode centrosomal protein and play a critical role in regulating proliferative divisions of neuroepithelial cells and neurogenesis, and participates in malignant progression of many kind of tumors(21, 22). The overexpression of ASPM had been proved to be related with the pathological staging and poor prognosis in liver cancer, pancreatic cancer, ovarian cancer and glioblastoma(23-25).

To explore potential underlying mechanism of the candidate hub genes, we constructed TF -miRNA-gene network of 21 hub genes, and the result shown CENPF had the most significant value in the network.

CENPF (centromere protein F), which mainly expressed in the G2/M cell cycle phase, is an important protein for cell cycle regulation by recruiting the checkpoint proteins like Mad1 and BubR1, resulting in a consistent response of checkpoint. CENPF has been identified as a key gene in several cancers especially in prostate cancer(26). CENPF have been reported as critical regulator of the COUP-TFII-FOXO1-CENPF axis in human prostate cancer, which was activated in the progress of metastasis and resulted in a poor prognosis in prostate cancer(27). Decrease the activity of CENPF can remarkable inhibit tumorigenesis in prostate cancer mouse models(28).

Even so ASPM and CENPF have been proved as a master regulator in several cancers, but there are rare reports about ASPM and CENPF associated with lung cancer, here we find that ASPM and CENPF are poor prognostic factors of lung cancer and both have high genetic alteration rate in the main pathological types of lung cancer. This result indicates that ASPM and CENPF may be potential therapy targets or biomarkers of lung cancer.

TF-miRNA-gene network revealed the most relevant TF are NFYA, E2F1 and MYC, and CENPF was the central node in the network, and the association between CENPF and these TFs are rarely been reported. E2F1 and MYC play important role in cell cycle progression and cell apoptosis and associated with the progress of several tumors(26, 29), these means E2F1 and MYC may interact with CENPF and participate in the tumorigenesis by regulating the cell cycle.

Conclusion

In this study we identified ASPM and CENPF as key genes in three main pathologic types of main pathological types of lung cancer, including LAC, LSCC and SCLC and these two genes may become potential therapy targets or biomarkers of lung cancer. At the same time, there are several limitations about our study. The main limitation is that the results of our study is obtained from microarray data and public databases through bioinformatic analysis, and can't provide real information on gene expression level, protein activity and genetic alteration, so multicenter and large sample follow-up studies are necessary to verify our results.

Materials And Methods

1.The Information of Microarray Data and Identification of DEGs

Lung cancer and normal lung tissues gene expression profile of GSE74706 including 10 LAC lung tissues, 8 LSCC lung tissues and 18 normal lung tissues, GSE60052 including 79 SCLC lung tissues and 7 normal control lung tissues, GSE19804 including 60 normal control lung tissues, 60 lung cancer tissues were obtained from NCBI-GEO. We analyzed the data according to the pathologic types of lung cancer. The data analysis of GSE74706 and GSE19804 was used GEOR2, an online tool of GEO. The data analysis of GSE60052 was used BioJupies (<https://amp.pharm.mssm.edu/biojupies/>) (12), an online

analysis tools of GEO RNA-seq Data. The criteria of DEGs was adjust P value $\leq$ 0.05 and $|\log FC| \geq 1$. We explored overlapped genes in different lung cancer pathologic types.

2.Functional Annotation and PPI Network of DEGs

To explore the function in a large scale of genes, the Gene Ontology (GO) term enrichment analysis and REACTOME pathway analysis of DEGs were performed by Enrichr (<http://amp.pharm.mssm.edu/Enrichr/>) (30), and the cut-off criterion was P value $\leq$ 0.05. The PPI network, which can visualize the interaction and reveal the relationship between each proteins, was constructed by The Search Tool for the Retrieval of Interacting Genes/Proteins (STRING) (<http://string-db.org>) (31) with the threshold of confidence score >0.4 . We then use Cytoscape software platform (32) to visualize and further analyze the PPI network.

3.Key Candidate Genes Screening

In the PPI network, each node has a connectivity value according to the strength of connection with neighbor nodes, we can calculate the degree of connectivity of each node by using an APP of Cytoscape software platform called cytoHubba, which can rank the nodes according to the properties of nodes by different algorithm, the node has a higher value means the node is more important in the network and may play a more important role in the biological process. To add credibility to the results, we used two kinds of topological algorithms (MCC, Degree) to calculate the top-scoring 25 nodes, and choose overlapped nodes in both algorithms method as hub genes for further study.

4.Kaplan-Meier Survival Analysis of the Candidate Genes

In order to explore the prognosis relationship between genes and LC, we performed Kaplan-Meier survival analysis of the hub genes by using Kaplan-Meier Plotter (<http://kmplot.com/analysis/>), which contain detailed survival data of 2437 lung cancer patients. The criteria was log-rank P value $\leq$ 0.05. The genes who had a significant association with the prognosis of LC were selected for further study.

5.The Genetic Alteration Analysis of Candidate Genes in LC patients

The cBioPortal for Cancer Genomics(<http://www.cbioportal.org/>) (13) is a powerful open-source resource and open-access TCGA dataset for investigation of multidimensional cancer genomics. This dataset can provide a high-quality information about molecular profiles, genetic alteration and clinical attributes from large-scale cancer genomics projects and helpful for researchers to translate the data into biologic insights and clinical applications. We choosed five study about LC totally including 3246 LC samples from the cBioPortal for Cancer Genomics dataset to explore the genetic alteration rate and co-occurrence relationship of the hub genes in LC patients.

6.Transcription Factor (TF) -miRNA-Gene Network of the Candidate Genes

Both TFs and miRNAs play critical roles in the regulation of gene and participate in the pathogenesis of cancer, the NetworkAnalyst ([https:// www. networkanalyst.ca/](https://www.networkanalyst.ca/)) (14) is a visual analytics platform for comprehensive gene expression profiling and meta-analysis, it provides comprehensive analysis for multi-kinds of PPI network construction as well as TF-miRNA-gene network construction. We use NetworkAnalyst to explore potential regulatory mechanism of these genes by constructing a TF-miRNA-gene network.

7.Function and Pathway Analysis of ASPM and CENPF

The PPI network of ASPM and CENPF the was constructed by using STRING to investigate the relationship of these two genes and explore the associated genes with these two genes with the threshold of confidence score >0.4. Gene Ontology analysis and KEGG pathway analysis of ASPM and CENPF associated genes were performed in STRING with the cut-off criterion P value $\leq$ 0.05.

Abbreviations

LC, lung cancer;

LSCC, Lung squamous cell carcinoma;

SCLC, small cell lung cancer;

LAC, Lung adenocarcinoma;

PPI, protein-protein interaction;

DEGs, differentially expressed genes;

TF, transcription factor.

Declarations

Acknowledgements

We wish to thank Dr. Chunlei zhou (Department of Pathology, Children's Hospital of Nanjing Medical University) for the manuscript review and literature support. We wish to thank Mr. Jing Zhang (Department of Cardiovascular Surgery, The First Affiliated Hospital of Nanjing Medical University) for the help with bioinformatics analysis.

Availability of data and materials

The datasets produced in current study is available from the corresponding author on reasonable request.

Authors' contributions

Sheng Zhao designed the study and coordination of the study. Jinghang Li analyzed the data and edited the manuscript. Jing Zhang help to analyzed the data and revise the manuscript. All authors read and approved the final manuscript.

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Competing interests

The authors declare that they have no competing interests.

References

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: a cancer journal for clinicians*. 2018;68(6):394-424.
2. Cohen AJ, Brauer M, Burnett R, Anderson HR, Frostad J, Estep K, et al. Estimates and 25-year trends of the global burden of disease attributable to ambient air pollution: an analysis of data from the Global Burden of Diseases Study 2015. *Lancet*. 2017;389(10082):1907-18.
3. Aran V, Victorino AP, Thuler LC, Ferreira CG. Colorectal Cancer: Epidemiology, Disease Mechanisms and Interventions to Reduce Onset and Mortality. *Clinical colorectal cancer*. 2016;15(3):195-203.
4. Frusch N, Bosquee L, Louis R. [Lung cancer. Epidemiology and etiologic factors]. *Revue medicale de Liege*. 2007;62(9):548-53.
5. Schwartz AG, Prysak GM, Bock CH, Cote ML. The molecular epidemiology of lung cancer. *Carcinogenesis*. 2007;28(3):507-18.
6. Fang S, Wang Z. EGFR mutations as a prognostic and predictive marker in non-small-cell lung cancer. *Drug design, development and therapy*. 2014;8:1595-611.
7. Sequist LV, Soria JC, Goldman JW, Wakelee HA, Gadgeel SM, Varga A, et al. Rociletinib in EGFR-mutated non-small-cell lung cancer. *The New England journal of medicine*. 2015;372(18):1700-9.
8. Cappuzzo F, Bemis L, Varella-Garcia M. HER2 mutation and response to trastuzumab therapy in non-small-cell lung cancer. *The New England journal of medicine*. 2006;354(24):2619-21.
9. Urabe F, Matsuzaki J, Yamamoto Y, Kimura T, Hara T, Ichikawa M, et al. Large-scale Circulating microRNA Profiling for the Liquid Biopsy of Prostate Cancer. *Clinical cancer research : an official journal of the American Association for Cancer Research*. 2019.

10. Gordian E, Welsh EA, Gimbrone N, Siegel EM, Shibata D, Creelan BC, et al. Transforming growth factor beta-induced epithelial-to-mesenchymal signature predicts metastasis-free survival in non-small cell lung cancer. *Oncotarget*. 2019;10(8):810-24.
11. Roth A, Boulay K, Gross M, Polycarpou-Schwarz M, Mallette FA, Regnier M, et al. Restoring LINC00673 expression triggers cellular senescence in lung cancer. *RNA biology*. 2018.
12. Torre D, Lachmann A, Ma'ayan A. BioJupies: Automated Generation of Interactive Notebooks for RNA-Seq Data Analysis in the Cloud. *Cell systems*. 2018;7(5):556-61.e3.
13. Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer discovery*. 2012;2(5):401-4.
14. Xia J, Gill EE, Hancock RE. NetworkAnalyst for statistical, visual and network-based meta-analysis of gene expression data. *Nature protocols*. 2015;10(6):823-44.
15. Liotta LA, Steeg PS, Stetler-Stevenson WG. Cancer metastasis and angiogenesis: an imbalance of positive and negative regulation. *Cell*. 1991;64(2):327-36.
16. Folkman J, . The role of angiogenesis in tumor growth. *Seminars in Oncology*. 2002;29(6):15-8.
17. Drabsch Y, ten Dijke P. TGF-beta signalling and its role in cancer progression and metastasis. *Cancer metastasis reviews*. 2012;31(3-4):553-68.
18. Papageorgis P. TGFbeta Signaling in Tumor Initiation, Epithelial-to-Mesenchymal Transition, and Metastasis. *Journal of oncology*. 2015;2015:587193.
19. Cicenias J, Kalyan K, Sorokinas A, Jatulyte A, Valiunas D, Kaupinis A, et al. Highlights of the Latest Advances in Research on CDK Inhibitors. *Cancers*. 2014;6(4):2224-42.
20. Berrier AL, Yamada KM. Cell-matrix adhesion. *Journal of cellular physiology*. 2007;213(3):565-73.
21. Zhong X, Liu L, Zhao A, Pfeifer GP, Xu X. The abnormal spindle-like, microcephaly-associated (ASPM) gene encodes a centrosomal protein. *Cell cycle (Georgetown, Tex)*. 2005;4(9):1227-9.
22. Bikeye SN, Colin C, Marie Y, Vampouille R, Ravassard P, Rousseau A, et al. ASPM-associated stem cell proliferation is involved in malignant progression of gliomas and constitutes an attractive therapeutic target. *Cancer cell international*. 2010;10:1.
23. Satow R, Shitashige M, Kanai Y, Takeshita F, Ojima H, Jigami T, et al. Combined functional genome survey of therapeutic targets for hepatocellular carcinoma. *Clinical cancer research : an official journal of the American Association for Cancer Research*. 2010;16(9):2518-28.
24. Wang WY, Hsu CC, Wang TY, Li CR, Hou YC, Chu JM, et al. A gene expression signature of epithelial tubulogenesis and a role for ASPM in pancreatic tumor progression. *Gastroenterology*. 2013;145(5):1110-20.
25. Horvath S, ., Zhang B, ., Carlson M, ., Lu KV, Zhu S, ., Felciano RM, et al. Analysis of oncogenic signaling networks in glioblastoma identifies ASPM as a molecular target. *Proceedings of the National Academy of Sciences of the United States of America*. 2006;103(46):17402-7.

26. Aytes A, Mitrofanova A, Lefebvre C, Alvarez MJ, Castillo-Martin M, Zheng T, et al. Cross-species regulatory network analysis identifies a synergistic interaction between FOXM1 and CENPF that drives prostate cancer malignancy. *Cancer cell*. 2014;25(5):638-51.
27. Lin SC, Kao CY, Lee HJ, Creighton CJ, Ittmann MM, Tsai SJ, et al. Dysregulation of miRNAs-COUP-TFII-FOXM1-CENPF axis contributes to the metastasis of prostate cancer. *Nature communications*. 2016;7:11418.
28. Mitrofanova A, Aytes A, Zou M, Shen MM, Abate-Shen C, Califano A. Predicting Drug Response in Human Prostate Cancer from Preclinical Analysis of In Vivo Mouse Models. *Cell reports*. 2015;12(12):2060-71.
29. Yu L, Fang F, Lu S, Li X, Yang Y, Wang Z. lncRNA-HIT promotes cell proliferation of non-small cell lung cancer by association with E2F1. *Cancer gene therapy*. 2017;24(5):221-6.
30. Chen EY, Tan CM, Kou Y, Duan Q, Wang Z, Meirelles GV, et al. Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC bioinformatics*. 2013;14:128.
31. Franceschini A, Szklarczyk D, Frankild S, Kuhn M, Simonovic M, Roth A, et al. STRING v9.1: protein-protein interaction networks, with increased coverage and integration. *Nucleic acids research*. 2013;41(Database issue):D808-15.
32. Smoot ME, Ono K, Ruscheinski J, Wang PL, Ideker T. Cytoscape 2.8: new features for data integration and network visualization. *Bioinformatics (Oxford, England)*. 2011;27(3):431-2.

Figures

Figure 1

Venn diagram of overlapped DEGs in GSE74706, GSE60052, GSE19804 dataset according to different pathologic types of lung cancer.

BP

- positive regulation of vasculature development (GO:1904018)
- regulation of angiogenesis (GO:0045765)
- positive regulation of angiogenesis (GO:0045766)
- venous blood vessel development (GO:0060841)
- negative regulation of sodium ion transmembrane transporter activity (GO:2000650)
- sprouting angiogenesis (GO:002040)
- regulated exocytosis (GO:0045055)
- positive regulation of nitric oxide biosynthetic process (GO:0045429)
- positive regulation of epithelial cell migration (GO:0010634)
- kidney development (GO:0001822)

CC

- condensed nuclear chromosome, centromeric region (GO:0000780)
- lamellar body (GO:0042599)
- integral component of plasma membrane (GO:0005887)
- condensed chromosome kinetochore (GO:0000777)
- condensed nuclear chromosome kinetochore (GO:0000778)
- spindle (GO:0005819)
- G-protein coupled receptor dimeric complex (GO:0038037)
- platelet alpha granule (GO:0031091)
- membrane raft (GO:0045121)
- condensed chromosome, centromeric region (GO:0000779)

MF

- transforming growth factor beta-activated receptor activity (GO:0005024)
- transmembrane receptor protein serine/threonine kinase activity (GO:0004675)
- protein homodimerization activity (GO:0042803)
- integrin binding (GO:0005178)
- transforming growth factor beta binding (GO:0050431)
- histone kinase activity (GO:0035173)
- amyloid-beta binding (GO:001540)
- BMP receptor activity (GO:0098821)
- endopeptidase inhibitor activity (GO:0004866)
- kinase binding (GO:0019900)

Reactome Pathway

- Cell Cycle, Mitotic_Homo sapiens_R-HSA-69278
- Cell Cycle_Homo sapiens_R-HSA-1640170
- Hemostasis_Homo sapiens_R-HSA-109582
- Platelet degranulation_Homo sapiens_R-HSA-114603
- Extracellular matrix organization_Homo sapiens_R-HSA-1474244
- Signaling by Rho GTPases_Homo sapiens_R-HSA-194315
- Response to elevated platelet cytosolic Ca2+_Homo sapiens_R-HSA-76005
- Mitotic Prometaphase_Homo sapiens_R-HSA-68877
- Resolution of Sister Chromatid Cohesion_Homo sapiens_R-HSA-2500257
- Platelet activation, signaling and aggregation_Homo sapiens_R-HSA-76002

Figure 2

Gene Ontology analysis and Reactome pathway analysis of DEGs. biological process (BP) group, molecular function (MF) group, cellular component (CC) group and reactome pathway was analyzed by Enrichr with $p < 0.05$ as the cut-off criterion. The length of the bar represents the gene number of enrichment.

Figure 3

PPI network of 459 DEGs. The green nodes represent DEGs, the lines between nodes represent the interaction of two nodes.

Figure 4

The hub genes identified from the PPI network. Hub genes was identified by CytoHubba using two kinds of topological algorithms (Degree, MCC) to calculate the top 25 nodes respectively. A. Top 25 hub genes in degree algorithms. B. Top 25 hub genes in MCC algorithms. C. the Venn diagram of 23 overlapping hub genes in both algorithms.

Figure 5

Kaplan–Meier analysis of 23 hub genes in the Kaplan–Meier plotter database about lung cancer patients. Only show the significant 21 hub genes. The red curve represents high expression of the gene. The dark curve represents low expression of the gene.

Figure 6

The genetic alteration analysis of Candidate Genes in 3191 lung cancer patients in the main pathological type of lung tumor. 19 genes have shown genetic alterations in lung cancer.

Figure 7

TF -miRNA-Gene Network of the Candidate Genes. Red nodes represent candidate genes, green square represent TF, blue square represent miRNA. The size of the nodes and squares means the significant degree in the network.

Figure 8

PPI network of ASPM.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Supplementarytables.pdf](#)