

RESEARCH ARTICLE

Identification of Driver Genes in Lung Squamous Cell Carcinoma and Lung Adenocarcinoma

Zhengtai Li¹, Fusheng Li², Huixue Tang¹, Xiao Ge¹, Lida Xu^{1*}, Changyuan Yu^{1*}

¹College of Life Science and Technology, Beijing University of Chemical Technology, Beijing, 100029, China

²SYSVAX INC., Beijing, 100176, China

Received: October 14, 2019

Accepted: November 25, 2019

Published Online: December 6, 2019

***CORRESPONDING AUTHORS**

Lida Xu,

Email: xuld@mail.buct.edu.cn

Changyuan Yu,

Email: yucy@mail.buct.edu.cn

CITATION

Li Z, Li F, Tang H, *et al.*, 2019, Identification of Driver Genes in Lung Squamous Cell Carcinoma and Lung Adenocarcinoma. *Cancer+*, 1(4):13-24.

DOI: 10.18063/cp.v1i4.258

Copyright: © 2019 Li, *et al.*

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), permitting all noncommercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract: The non-small cell lung cancer (NSCLC), including lung squamous cell carcinoma (LUSC) and lung adenocarcinoma (LUAD), accounts for a large proportion of lung cancer cases. However, the mechanisms of LUSC and LUAD are very different, especially the pathogenesis of LUSC remains unclear. At present, the research on the targeted therapeutic sites of LUAD has approached maturity and these targets are of clinical significance. However, effective therapeutic targets have not been identified in LUSC, and at present, the same targeted therapeutic strategy for LUAD is also applied in LUSC treatment. We used the data from The Cancer Genome Atlas program to analyze the driver genes of LUAD and LUSC by two types of algorithms, namely, the OncodriveCLUST and Multi-Dendrix. Our results showed that the driver genes of LUAD concentrates in the KRAS/epidermal growth factor receptor/TP53 pathways, while LUSC involves multiple pathways, including PIK3CA, NFE2L2, and TP53. The results showed that different carcinogenic mechanisms exist between these two types of NSCLC, implying that different therapeutic targets for LUSC deserve our attention. At the same time, the results of survival analysis proved that the driver genes identified using the two algorithms in combination may be more valid and reliable than those identified by solely using MutsigCV.

Keywords: Driver genes, Non-small cell lung cancer, OncodriveCLUST algorithm, Multi-Dendrix algorithm

1 Introduction

Lung cancer has become one of the most common cancers that threaten human health^[1]. Currently, according to the morphological standard, lung cancer is mainly classified into small cell lung cancer and non-small cell lung cancer (NSCLC). About 85% of lung cancers can be classified as NSCLC^[2], among which lung adenocarcinoma (LUAD) and lung squamous cell carcinoma (LUSC) occupy a considerable proportion. According to the clinical treatment guidelines^[3], targeted therapies for NSCLC with various mutated genes such as KRAS, epidermal growth factor receptor (EGFR), anaplastic lymphoma kinase (ALK), BRAF, and ROS1 are available. However, these targeted therapeutic sites display extremely low mutation frequency in LUSC. The KRAS mutations have been reported frequently in

non-smoking patients with LUAD^[4,5], but these mutations are quite rare in LUSC^[6]. Furthermore, the frequencies of mutations at many key pathogenic genes show great differences between LUSC and LUAD^[7].

Genes that play important roles in the development and progression of tumors are called driver genes. The driver genes generally occupy and control the critical pathway of a major physiological process and may cause disorder in the downstream pathways^[8]. Therefore, driver genes can be used as targets for tumor therapy^[9]. On the contrary, the genes which are subject to mutations as a consequence of the mutations to the driver genes are usually called passenger genes. Compared with the driver genes, passenger genes mutate in the course of tumor development following the changes to the driver genes. Thus, the passenger genes that also display a relatively high mutation frequency and a variety of mutation types are difficult to be distinguished from the driver genes. On the other hand, the driver genes do not always display high mutation frequencies in tumor. For example, Buisson *et al.* found that the most prominent cancer driver genes on the chromosome 17 such as TP53, NF1, ERBB2 (Her2), and BRCA1 are all located in regions with the lowest frequency of background mutation^[10]. In view of this, mining the driver genes of tumors is complicated and how to determine the driver genes will be the key problem.

Unlike the driver genes of LUAD which concentrate on several discovered sites or pathways^[11], the pathogenesis of LUSC is more complicated and multiple pathways are involved^[12]. By analyzing methylation data, the pathogenic pathways of these two types of NSCLCs are mainly derived from the pathways pertaining to fat digestion and absorption, phenylalanine metabolism, bile secretion, and other related nutrient metabolism pathways^[2]. However, more studies have shown that the driver genes of LUSC are mainly concentrated in pathways such as PIK3CA, TP53, PTEN, and SOX2^[6]. By comparing the pathogenic genes, immunotherapy targets, and many other aspects of LUAD and LUSC coupled with a large number of experimental data and literature, Relli *et al.*

found that LUAD and LUSC are very different at molecular, pathological, and clinical levels^[13]. Recently, targeted anti-vascular endothelial growth factor therapy with bevacizumab was reported to improve the survival of LUAD patients, whereas it was contraindicated in patients with LUSC because of the fatal hemoptysis^[3]. Above all, the occurring mechanism of LUAD is much different from that of LUSC, and the treatment strategies of LUAD maybe not ideal to LUSC. Therefore, it is very important to identify the pathogenesis of LUSC and the new potential therapeutic targets.

Discovering driver genes by only detecting the frequencies of gene mutations have great limitations. This approach does not take the intrinsic relationship of mutated genes into account, so the driver genes with relatively lower mutation frequencies can be easily ignored or the passenger genes can be picked out in the process^[6,13]. Some algorithms have been developed based on the hypothesis that gain-of-function mutations often cluster in specific protein regions. For example, the OncodriveFM algorithm^[14] has been applied to identify genes with driving variability in tumor development, by evaluating the differential accumulation of mutated genes with high functional impact. Such methods have the potential to unearth some genes that are not high in frequency but important in function. Another algorithm OncodriveCLUST evaluates the biased degree of the genes toward mutation clusters, which formed by considering the functions of genes combined with their variation frequencies. In comparison of these two cluster-based detecting algorithms, OncodriveCLUST is more likely to find out reasonable driver gene candidates than OncodriveFM. Moreover, OncodriveCLUST has been integrated into the R package “maftools”^[15], making it easier to be conducted.

Since cancer is considered to be caused by the accumulative dysfunction of multiple genes in regulating particular signal pathways, another idea for identifying cancer driver genes is to discover the driver pathways instead. Vaske *et al.* proposed a method for identifying cancer-driven pathways based on existing genomics data and

gene expression data using the factor graph model^[16]. In recent years, a series of studies on inferring driver genes by combining a plurality of networks with cancer-related genes or RNA expression data have obtained many meaningful results. The Dendrix algorithm^[17], one of these methods, transforms the problem of path identification into solving the maximum weight submatrix with Markov Chain Monte Carlo method. The parameter k in the algorithm is the number of gene sets with high coverage and high mutual exclusion. In addition, the genes within the sets will be selected as the driver genes. To further adapt to analyzing tumor driver genes in multiple pathways, a multi-path algorithm Multi-Dendrix^[18] was then developed to solve this problem and improve computational efficiency.

MutSigCV is one of the commonly used algorithms for driver genes detection and it is recommended in The Cancer Genome Atlas (TCGA) website. It does not only consider the gene mutation frequency in samples but also the mutation rate compared with the background expression data and so on^[19]. Compared to MutsigCV, the OncodriveCLUST and Multi-Dendrix mainly focus on gene functions with the previous knowledge in some biological databases. In this paper, we chose to identify the driver genes using OncodriveCLUST and Multi-dendrix, and then validated and discussed the results by comparing to the candidates found with MutSigCV.

In this study, the driver genes of LUSC were mined using these two types of algorithms to find out comprehensive potential therapeutic targets. The differences in pathogenesis between these two typical subtypes of NSCLCs were then discussed by comparing their tumor-related pathways. Furthermore, the functional bias of the two driver genes mining algorithms was compared and discussed.

2 Materials and methods

The Mutation Annotation Format files of samples were downloaded from TCGA program website (<https://portal.gdc.cancer.gov/>). The OncodriveCLUST method in R package “maftools” was used to analyze the driver genes

of LUSC and LUAD (<http://bioconductor.org/packages/release/bioc/html/maftools.html>). The minimum frequency of gene mutations to be selected was set to 15, and the clustering method was using “z-score.” To compare the results generated by different types of algorithms, LUSC and LUAD driver genes were also identified using the Multi-Dendrix package (<http://compbio.cs.brown.edu/multi-dendrix/>). The number of gene sets was limited to between 2 and 4 according to the recommended range provided by the software, and the size of the gene sets was limited to between 3 and 5^[17]. The results of the two methods were enriched in KEGG pathway analysis to identify the key pathogenesis pathways. MutsigCV with its recommended settings was used in the final step to validate the results.

3 Results

3.1 Baseline characteristics

Samples of 491 LUSC and 560 LUAD from TCGA were analyzed in this study. **Figure 1** shows the distribution of these samples in terms of smoking history and gender. Male patients accounted for nearly three quarters of the LUSC cases, and smokers for more than 80% in the samples^[20]. Since the sample ratio was comparable to the previous reports^[6], the results from these samples could be extrapolated to the entire population to a certain extent.

3.2 Frequencies of gene mutations in LUSC and LUAD

The 15 genes with the highest mutation frequencies in the LUSC and LUAD datasets are displayed in **Figure 2**. TP53, which is a common tumor suppressor gene and always has a very high mutation frequency in various cancers, had the highest mutation rates; the mutation rates were recorded at 79% in LUSC cases (**Figure 2A**) and 47% in LUAD cases (**Figure 2B**). The ten genes, including TTN, MUC16, RYR2, CSMD3, LRP1B, ZFHX4, USH2A, XIRP2, SPTA1, and NAV3, appeared in both mutant gene sets of LUSC and LUAD. However, MUC16 and TTN genes with high mutation frequencies have not

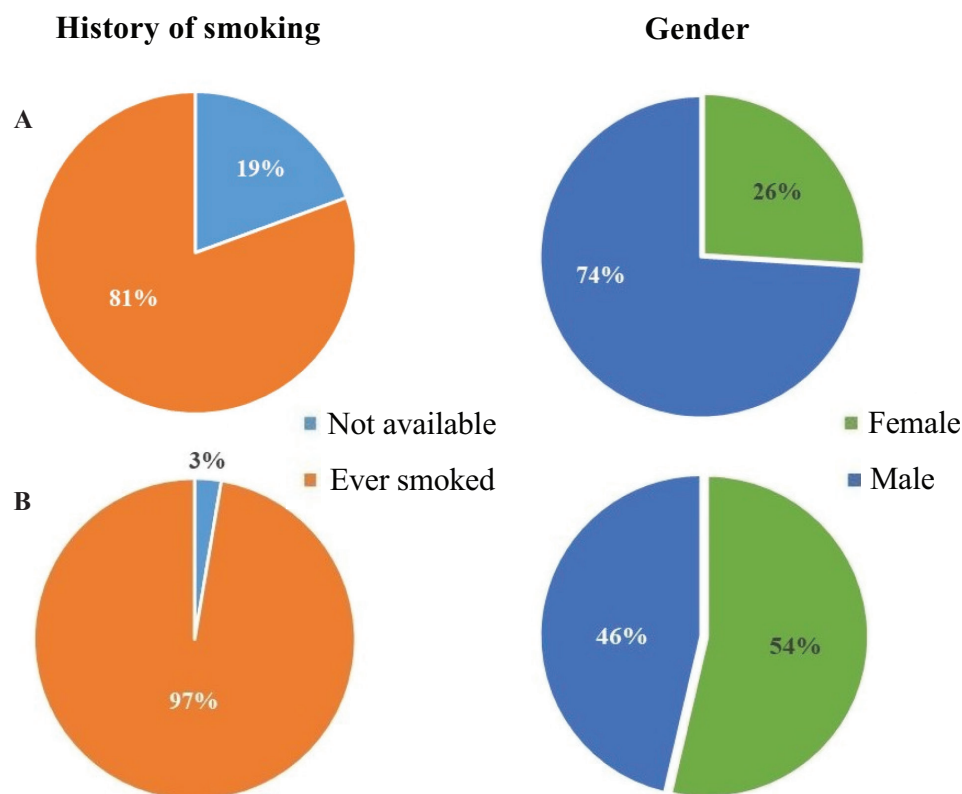


Figure 1. The proportion of smoking history and gender in lung squamous small cell carcinoma (LUSC) and lung adenocarcinoma (LUAD) samples. The history of smoking and gender distribution in LUSC and LUAD samples is shown in panels (A) and (B), respectively.

shown strong driving effects on the development of LUSC in the previous studies^[3]. Based on the statistics of single-nucleotide variant class, the rate of C > T mutations in LUSC (**Figure 2C**) is somewhat greater than that in LUAD (**Figure 2D**). It has also been reported in an early study that a little portion of LUSC patients presented a type of high-frequent C > T mutation signature, which was obviously different from LUAD patients^[21].

3.3 Driver genes of LUSC

Ten genes which were identified by the OncodriveCLUST algorithm have high potential in playing a key role in the development of LUSC. As shown in **Figure 3A**, the OncodriveCLUST analysis showed that the driver genes of LUSC, PIK3CA, PTEN, TP53, and NFE2L2 were preferentially highlighted with significant false discovery rate (FDR) values. However, certain genes such as MB21D2, LRRIQ3, MSR1, and WWC3 with the lower FDR values in the midstream stage have

never been reported on whether they could lead to LUSC directly or indirectly.

However, the driver genes obtained using the pathway recognition algorithm Multi-Dendrix were not the same as the results of the clustering algorithm. We screened the results of Multi-Dendrix based on the degree of correlation between genes and previously reported sites which may lead to LUSC. As shown in **Figure 3B**, four relatively independent networks were formed from the results of Multi-Dendrix, which indicated that the development of LUSC had a combination of multiple pathogenic factors. Overall, the pathway that controls the tumor suppressor function, involving the well-known tumor suppressor gene TP53, leads to tumor cell proliferation. The second group mainly controls the pathway of the calcium ions, such as the MMP16, PCDHB14, LRP2, and CDCNA1A (**Figure 3B**). The last group is involved in the tumor cell production pathway (**Figure 3B**)

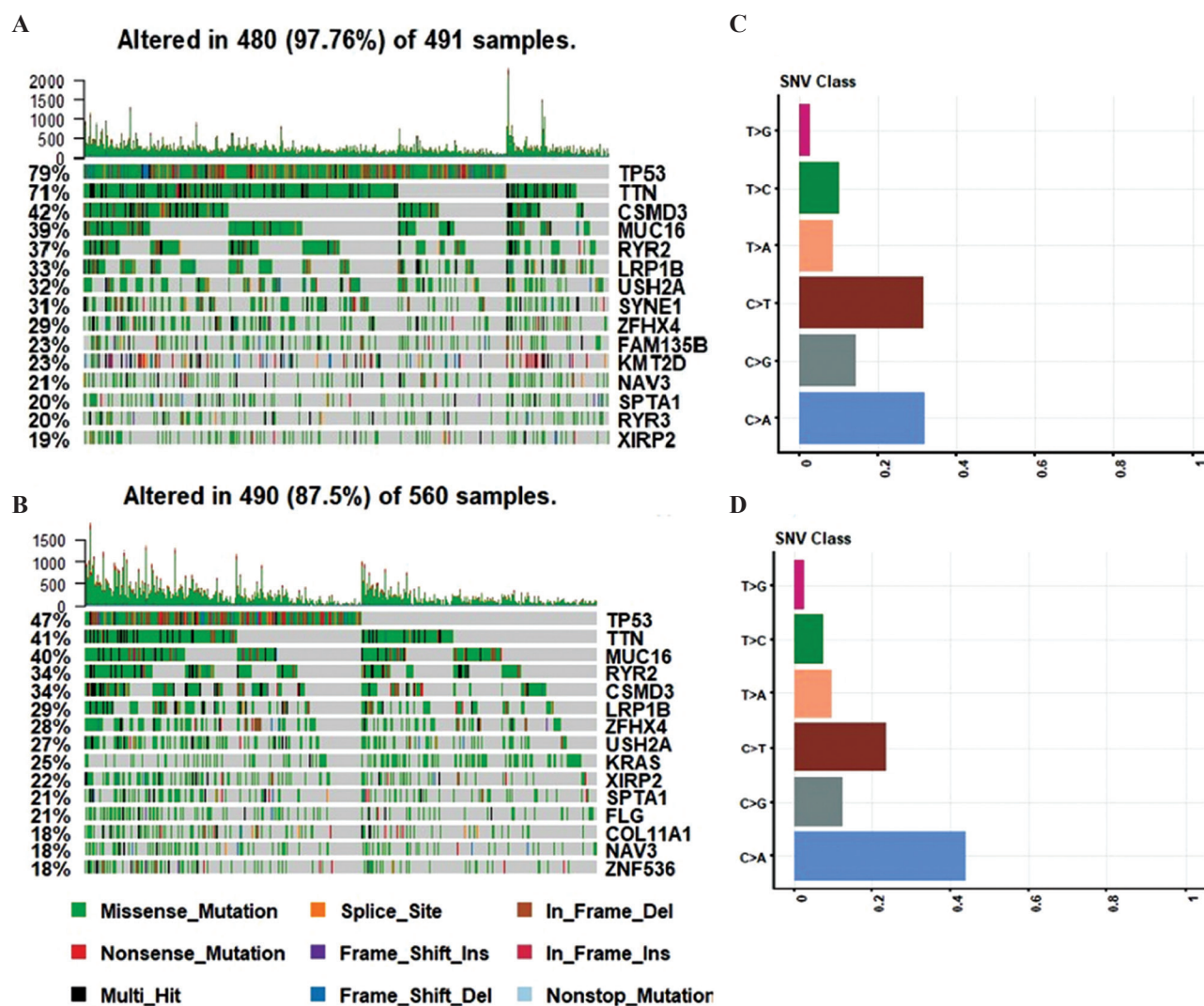


Figure 2. The top 15 genes with the highest mutation frequencies and the frequencies of single-nucleotide variant (SNV) class in lung squamous cell carcinoma (LUSC) and lung adenocarcinoma (LUAD). The mutation frequencies of 15 genes in LUSC and LUAD are shown in panels (A) and (B), respectively. The SNV class in LUSC and LUAD are shown in panels (C) and (D), respectively.

controlled by NFE2L2, which encodes protein Nrf2^[22]. Nrf2, a redox-sensitive transcription factor, regulates the expression of antioxidant enzymes and several anti-apoptotic proteins, and it helps to confer cytoprotection against oxidative stress and apoptosis^[22]. LRP1B is also a commonly mutated gene in NSCLC^[5].

Combining the results of the two algorithms, 16 potential driver genes of LUSC were obtained.

3.4 LUSC and LUAD have different driver genes

To explore the differences in driver genes LUAD and LUSC, analysis of driver genes of LUAD was conducted in the same way.

According to the OncodriveCLUST results on LUAD, the driver genes include those with higher mutation frequency and with lower mutation frequency, for example, the EGFR and CRISP2 that were mutated in 12.41% and 1.79% of the samples, respectively. KRAS along with EGFR, the current mainstream therapeutic target, took the dominant position (**Figure 4A**), although they did not display high mutation frequencies. At the same time, MUC16 that was used as a prognostic target for many cancers in previous studies, was also found as an important driver gene.

Using the Multi-Dendrix algorithm, certain genes such as KRAS, EGFR, TP53, and MUC16

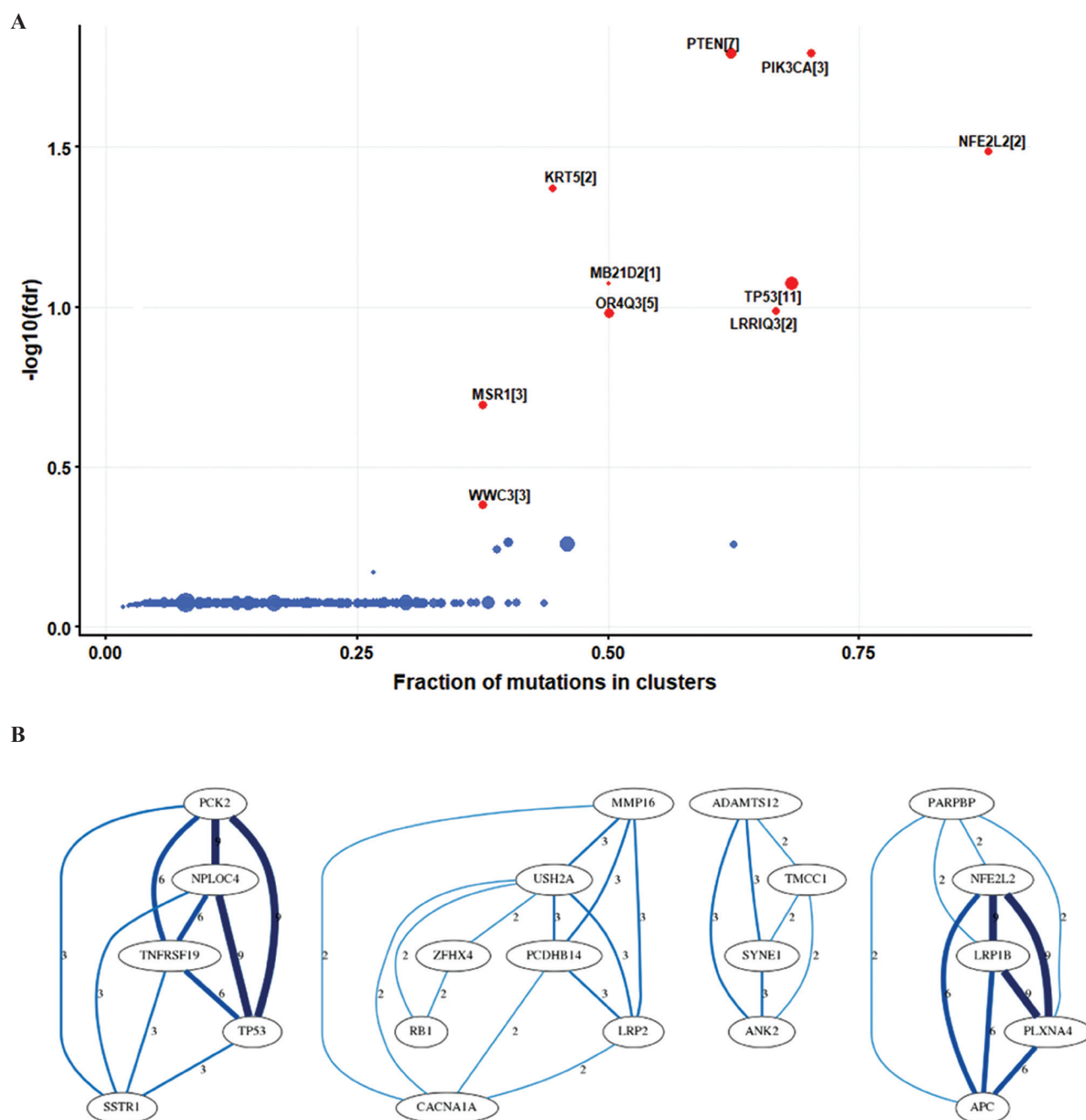


Figure 3. Driver genes of lung squamous cell carcinoma (LUSC) identified by the OncodriveCLUST algorithm and Multi-dendrix algorithm. The driver genes of LUSC which were identified by OncodriveCLUST algorithm were shown in panel (A) in which the vertical axis represents the false discovery rate, the horizontal axis represents the proportion of clustering to the point, and the number in square brackets indicates the number of clusters involving the gene. The driver genes of LUSC which were identified by Multi-dendrix algorithm were shown in panel (B) in which the numbers represent the frequencies of the gene pairs appearing in the same gene set.

were found to be the important driver genes appeared (**Figure 4B**). Moreover, ERBB2 was also in a dominant position, which is in agreement with current findings.

3.5 Comparison of the candidate driver genes obtained from MutsigCV

MutsigCV is widely used to seek the driver genes. While the above two algorithms focused

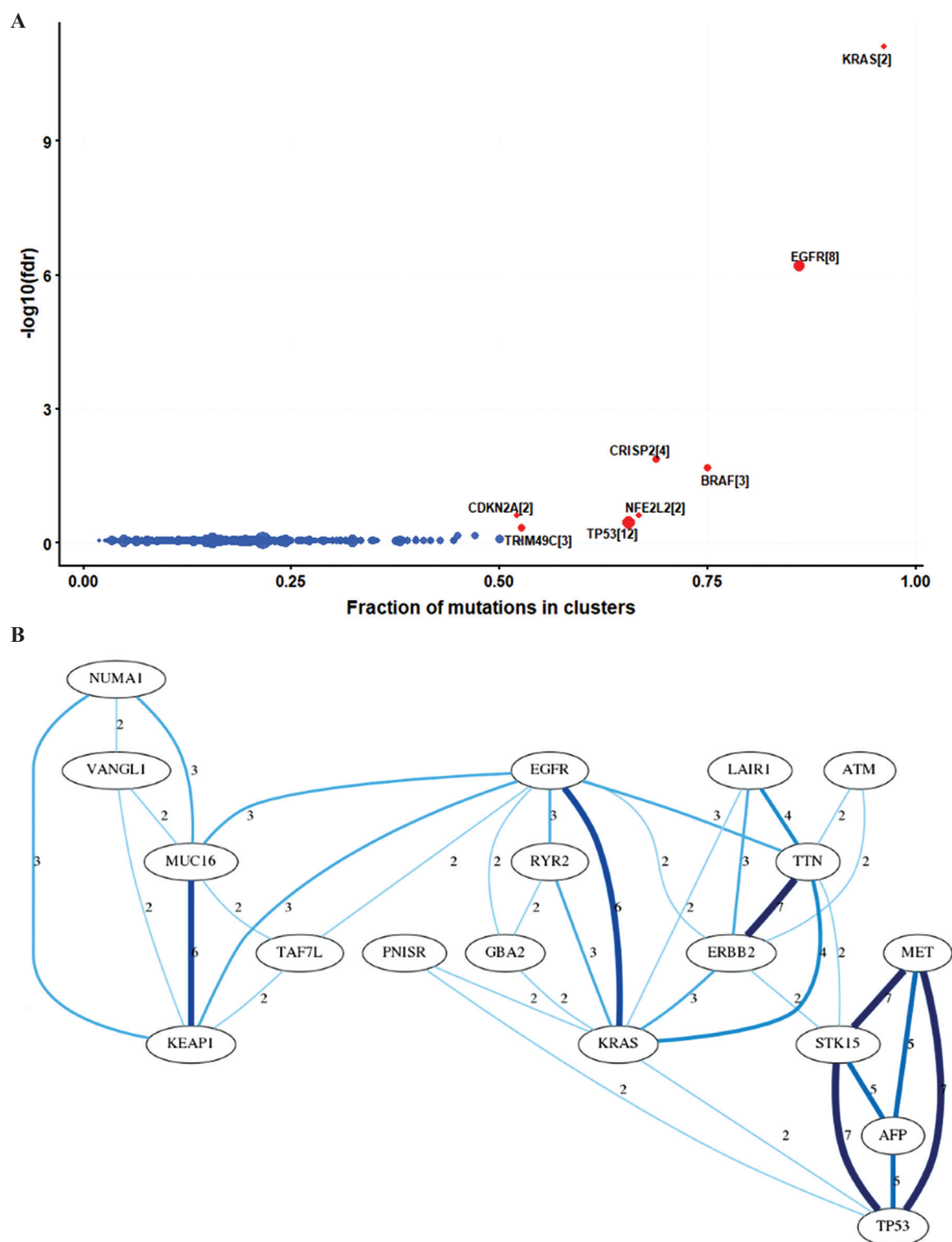


Figure 4. Driver genes of lung adenocarcinoma (LUAD) identified by the OncodriveCLUST algorithm and Multi-dendrix algorithm. The driver genes of LUAD which were identified by OncodriveCLUST algorithm were shown in panel (A) in which the vertical axis represents the false discovery rate, the horizontal axis represents the proportion of clustering to the point, and the number in square brackets indicates the number of clusters involving the gene. The driver genes of LUAD which were identified by Multi-dendrix algorithm were shown in panel (B) in which the numbers represent the frequencies of the gene pairs appearing in the same gene set.

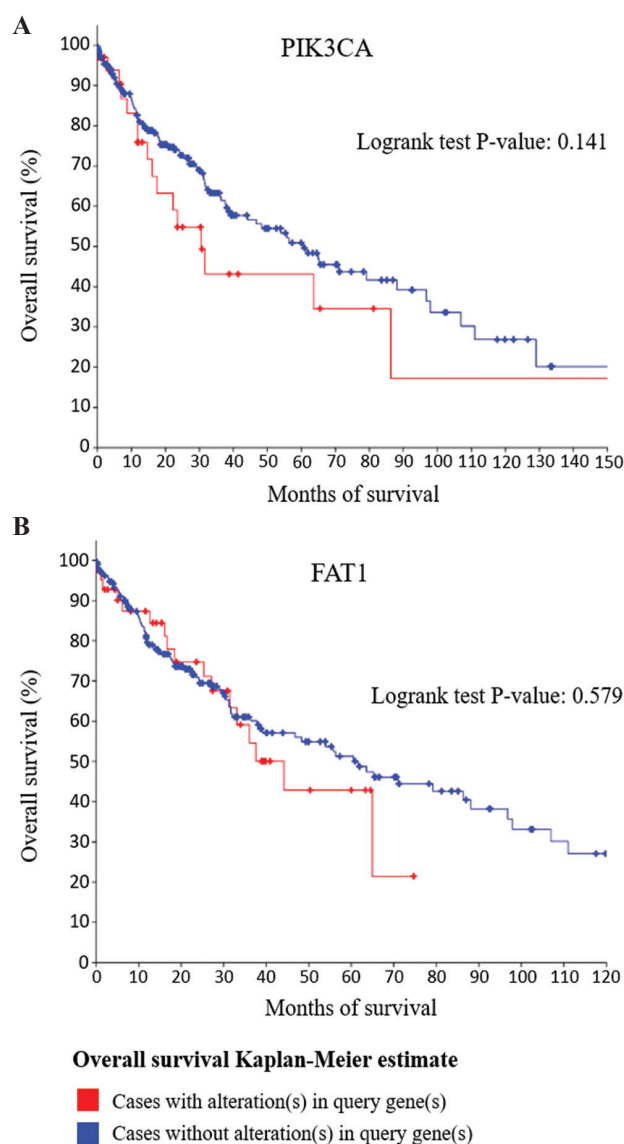


Figure 5. Survival analyses of (A) PIK3CA and (B) FAT1 mutations in lung squamous cell carcinoma (LUSC) samples without considering TP53 mutations.

on the function of mutated genes, MutsigCV evaluated the effects of the frequencies and types of gene mutations on gene expression or other phenotypes. Here, MutsigCV was adopted for comparison with the two algorithms, to gain a better insight into the results.

There were 48 genes which were significant (FDR-adjusted $P < 0.05$) in LUSC with MutsigCV. The top ten genes with the smallest FDR-adjusted P-values are presented in **Table 1**. While some of the candidate genes, including NFE2L2, TP53, and CDKN2A, obtained in

previous analysis were also significant in LUSC samples, but the OncodriveCLUST algorithm and Multi-dendrix algorithm as described above could not identify other candidates. According to MutsigCV analysis, more genes such as FAT1, ARID1A, and SLCO1B1 which have significant effects and had not been proven to have a significant impact on LUSC in previous studies were found. Meanwhile, PIK3CA that was in a dominant position in the previous results was not able to be selected in the results of MutsigCV.

To verify the results, survival analyses were conducted about these candidate driver genes. Because TP53 has a great impact on the survival time, cases with TP53 mutations were excluded from the study. The results presented that, in comparison to the unique significant genes determined in MutsigCV, the unique driver genes identified with OncodriveCLUST and Multi-Dendrix had more impact on the survival time. The survival analyses of PIK3CA (**Figure 5A**) and FAT1 (**Figure 5B**) mutations in LUSC showed that PIK3CA affected the survival of patients in a more obvious way.

Meanwhile in LUAD, MutsigCV derived three candidate genes including KRAS, EGFR and TP53 (**Table 2**), which were also discovered by OncodriveCLUST and Multi-Dendrix.

3.6 Driver pathways of LUSC and LUAD

The function of the 16 candidate genes associated with LUSC was analyzed in the KEGG pathway analysis. The PI3K-related pathway controlled by the PIK3CA gene was significant (**Figure 6A**). The PI3K pathway was also found to have strong synergy with the Nrf2 pathway in the development of lung cancer in the previous studies^[23].

In contrast, the detected driver genes of LUAD were more concentrated and enriched in the pathogenesis pathways of NSCLC in KEGG (**Figure 6B**).

4 Discussion

We used two different types of algorithms, namely, the OncodriveCLUST and Multi-Dendrix, to study the driver genes of LUSC using the TCGA data. To clarify the similarities and differences in the pathogenic mechanisms of

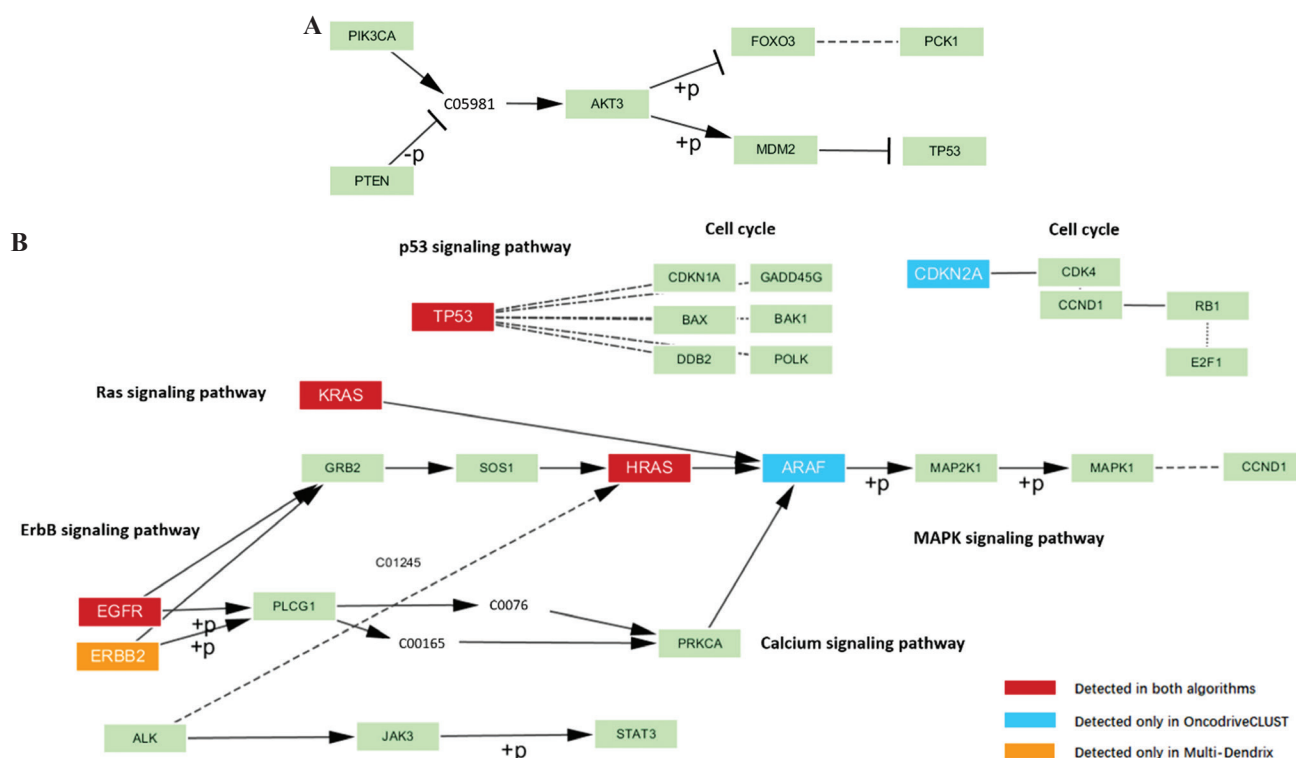


Figure 6. Driver pathways of lung squamous cell carcinoma (LUSC) and lung adenocarcinoma (LUAD). (A) Candidate driver genes associated with LUSC were enriched in the PI3K pathway using KEGG pathway analysis. (B) KEGG enrichment results of driver genes in LUAD.

Table 1. Driver genes of lung squamous cell carcinoma obtained from MutsigCV

Gene	<i>P</i>	FDR-adjusted <i>P</i>
CDKN2A	0	0
NFE2L2	0	0
RB1	9.99201E-16	6.28231E-12
PTEN	1.33227E-15	6.28231E-12
TP53	3.77476E-15	1.42399E-11
KEAP1	5.10703E-15	1.60548E-11
FAT1	1.79828E-08	4.23989E-05
ARID1A	7.19699E-08	0.000150833
NF1	3.09118E-06	0.005619934
SLCO1B1	1.21386E-05	0.01784985

FDR, false discovery rate.

Table 2. Driver genes of lung adenocarcinoma obtained from MutsigCV

Gene	<i>P</i>	FDR-adjusted <i>P</i>
FAM168A	0	0
KEAP1	0	0
KRAS	0	0
RB1	0	0
SMAD4	0	0
SMARCA4	0	0
STK11	0	0
TP53	1.77636E-15	3.72285E-12
EGFR	2.9643E-14	5.08296E-11

FDR, false discovery rate.

the two types of NSCLC, we also analyzed the driver genes of LUAD using the same methods.

The driver genes of LUAD identified by the two types of algorithms are relatively consistent and enriched in the KEGG pathways of NSCLC. KRAS, EGFR, and TP53 stood out at the key

positions of an upstream pathway and were considered the principal genes that cause LUAD, which is in agreement with the previous reports. Furthermore, these driver genes were discovered with the MutsigCV algorithm. In this case, both algorithms could identify the driver genes accurately.

However, while detecting the driver genes of LUSC, the results obtained by OncodriveCLUST and Multi-Dendrix algorithms have obvious differences. Thus, we integrated the mutation sites highlighted by the two algorithms. It can be seen that the mechanism of LUSC is more complicated since this form of NSCLC involves multiple pathways. First, PIK3CA has the most obvious effect on LUSC. PIK3CA that controls the activities of the transferase and the serine and threonine proteins leads to the production of LUSC as usually reported. The PTEN is also involved in this pathway. Using the multi-pathway algorithm, we found that there are two key pathways contributing to LUSC development. The two pathways include the pathway that inhibits tumor cell proliferation by TP53, and another controlled by NFE2L2. NFE2L2 gene encodes the Nrf2 protein, a redox-sensitive transcription factors that regulate the expression of antioxidant enzymes and several anti-apoptotic proteins and helps to confer cytoprotection against oxidative stress and apoptosis. PIK3CA and NFE2L2 often have strong synergistic effects in tumors and have been reported in the previous pathway studies for multiple tumors. However, it is worthy to notice that PIK3CA gene was not identified by MutsigCV. The second pathway in **Figure 3B** is related to the calcium channel controlled by the MMP16, which is strongly related to calcium ion binding (signaling pathway). However, there were no studies that support the role of ion channels in LUSC development. Calcium ions affect many critical physiological processes in humans to varying degrees, its relationship with tumor deserves further investigation and discussion.

In the present study, three genes (PIK3CA, TP53, and NFE2L2) and two pathways (PI3K and TP53) were found to have a great impact on LUSC development. At present, NFE2L2 is viewed as a therapeutic target in the clinic. For example, a typical NFE2L2 inhibitor ML385 has been proven to be effective against NSCLC^[24]. At the same time, we also analyzed the driver genes in LUAD which belongs to the family of NSCLC along with LUSC and found that their driver gene compositions have a huge difference. The driver

genes of LUAD are mainly KRAS and EGFR, which are consistent with the previous studies.

The two different types of algorithms, OncodriveCLUST and Multi-Dendrix, have different functional biases. The Multi-Dendrix algorithm is gene set-oriented at the beginning of the model establishment, giving priority to looking for genome files with high coverage, and then looking for subsequent gene sets to form the whole pathways related to tumor generation. It is clear that within the network diagram of known gene sets, genes have obvious continuity in their biological functions and have functional continuity between them. The NFE2L2 pathway controls the changes in upstream gene transcription; LRP1B and APC genes that have the function of suppressing tumor and finally affecting DNA transcription are associated with the NFE2L2 pathway.

The typical clustering algorithm represented by OncodriveCLUST tends to classify the mutated genes according to their functions and prioritize the background of the dataset. It also screens the mutation frequency of each gene and then analyzes the driver genes and examines if the functions of selected driver genes are the same. If the genes share parallel relationships, their functions would be similar.

In the case of LUAD, both algorithms have identified three important lung cancer-driving nodes, namely, KRAS, EGFR, and TP53, but none of the EML4-ALK fusion genes are included in the study. The reason is that the variation type may not be covered by the algorithm, and EML4-ALK fusion mainly happens within Asians who account for only a small proportion in the samples from the TCGA database. The difference in the results of the two algorithms is that the Dendrix algorithm uncovers the important role of ERBB2 coupled with EGFR in controlling the downstream pathways, as shown in **Figure 6B**. However, the pathway controlled by CDKN2A, a relatively principal tumor suppressor gene, was not found, and the mutation of BRAF, which is important in downstream pathways, was also ignored by the Dendrix algorithm.

In comparison with the results of the MutsigCV algorithm, it is not difficult to see that

the combination of OncodriveCLUST and Multi-Dendrix is more practical in the identification of driver genes than the sole application of MutsigCV algorithm that relies on the background mutation frequency. The use of combined algorithms can effectively reduce the possibility of finding false-positive results.

6 Conclusion

The driver genes of LUAD concentrate in the KRAS/EGFR/TP53 pathways, while the development of LUSC may involve multiple pathways, including PIK3CA, NFE2L2, and TP53. In addition, using the clustering algorithm OncodriveCLUST and pathway algorithm Multi-Dendrix can identify potential driver genes with higher accuracy.

Acknowledgments

This work was supported by the National Science and Technology Major Project during the 13th 5-Year Plan Period (project number: 2019ZX09721001-007-002) and Shenzhen Science and Technology Project (project number: JCYJ20180507183842516).

Conflict of interest

The authors declared that they have no conflicts of interest in this work and do not have any commercial or associative interest that represents a conflict of interest in connection with the work submitted.

Author contributions

Z.L., L.X., and C.Y. conceived and designed the experiments. L.X. and F.L. verified the analytical methods. Z.L., H.T., and X.G. analyzed the data. L.X. and F.L. conceived the main ideas and review outline. Z.L. and L.X. wrote the paper. L.X. and C.Y. reviewed drafts of the paper. All authors read and approved the final manuscript.

References

1. Siegel RL, Miller KD, Jemal A, 2018, Cancer Statistics, 2018. *CA Cancer J Clin.* 68(1):7–30.
2. Herbst RS, Morgensztern D, Boshoff C, 2018, The Biology and Management of Non-small Cell Lung Cancer. *Nature*, 553(7689):446–54. DOI: 10.1038/nature25183.
3. Lindeman NI, Cagle PT, Aisner DL, et al., 2018, Updated Molecular Testing Guideline for the Selection of Lung Cancer Patients for Treatment With Targeted Tyrosine Kinase Inhibitors: Guideline From the College of American Pathologists, the International Association for the Study of Lung Cancer, and the Association for Molecular Pathology. *J Thorac Oncol*, 13(3):323–58. DOI: 10.1097/jto.0000000000000002.
4. Pao W, Wang TY, Riely GJ, et al., 2005, KRAS Mutations and Primary Resistance of Lung Adenocarcinomas to Gefitinib or Erlotinib. *PLoS Med*, 2(1):e17. DOI: 10.1371/journal.pmed.0020017.
5. Ding L, Getz G, Wheeler DA, et al., 2008, Somatic Mutations Affect Key Pathways in Lung Adenocarcinoma. *Nature*, 455(7216):1069–75.
6. Jamal-Hanjani M, Wilson GA, McGranahan N, et al., 2017, Tracking the Evolution of Non-small-cell Lung Cancer. *N Engl J Med*, 376(22):2109–21.
7. Campbell JD, Alexandrov A, Kim J, et al., 2016, Distinct Patterns of Somatic Genome Alterations in Lung Adenocarcinomas and Squamous Cell Carcinomas. *Nat Genet*, 48(6):607–16.
8. Greenman C, Stephens P, Smith R, et al., 2007, Patterns of Somatic Mutation in Human Cancer Genomes. *Nature*, 6(9):153–8.
9. Zheng CH, Yang W, Chong YW, et al., 2016, Identification of Mutated Driver Pathways in Cancer Using a Multi-objective Optimization Model. *Comput Biol Med*, 72:22–9.
10. Buisson R, Langenbucher A, Bowen D, et al., 2019, Passenger Hotspot Mutations in Cancer Driven by APOBEC3A and Mesoscale Genomic Features. *Science*, 364(6447):1228. DOI: 10.1126/science.aaw2872.
11. Cancer Genome Atlas Research Network, 2014, Comprehensive Molecular Profiling of Lung Adenocarcinoma. *Nature*, 511(7511):543–50. DOI: 10.1038/nature13385.
12. Cancer Genome Atlas Research Network, 2012, Comprehensive Genomic Characterization of Squamous Cell Lung Cancers. *Nature*, 489(7417):519–25. DOI: 10.1038/nature11404.
13. Relli V, Trerotola M, Guerra E, et al., 2019, Abandoning the Notion of non-small Cell Lung Cancer. *Trends Mol Med*, 25(7):585–94. DOI: 10.1016/j.molmed.2019.04.012.
14. Gonzalez-Perez A, Lopez-Bigas N, 2012, Functional Impact Bias Reveals Cancer Drivers. *Nucleic Acids Res*, 40(21):e169. DOI: 10.1093/nar/gks743.
15. Mayakonda A, Lin DC, Assenov Y, et al., 2018, Maftools: Efficient and Comprehensive Analysis of Somatic Variants

- in Cancer. *Genome Res*, 28(11):1747-56. DOI: 10.1101/gr.239244.118.
16. Vaske CJ, Benz SC, Sanborn JZ, *et al.*, 2010, Inference of Patient-specific Pathway Activities from Multi-dimensional Cancer Genomics Data Using PARADIGM. *Bioinformatics*, 26(12):i237–45. DOI: 10.1093/bioinformatics/btq182.
17. Vandin F, Upfal E, Raphael BJ, 2012, *De novo* Discovery of Mutated Driver Pathways in Cancer. *Genome Res*, 22(2):375–85. DOI: 10.1101/gr.120477.111.
18. Leiserson MD, Blokh D, Sharan R, *et al.*, 2013, Simultaneous Identification of Multiple Driver Pathways in Cancer. *PLoS Comput Biol*, 9(5):e1003054. DOI: 10.1371/journal.pcbi.1003054.
19. Lawrence MS, Stojanov P, Polak P, *et al.*, 2013, Mutational Heterogeneity in Cancer and the Search for New Cancer-associated Genes. *Nature*, 499(7457):214–8.
20. Ahmad D, 2017, Epidemiology Study on P53 (Rs1614984) C>T Mutation in Cigarette Smokers. *Braz Arch Biol Technol*, 60:e160438. DOI: 10.1590/1678-4324-2017160438.
21. Alexandrov LB, Nik-Zainal S, Wedge DC, *et al.*, 2013, Signatures of Mutational Processes in Human Cancer. *Nature*, 500(7463):415–21.
22. Anju S, Manish B, Nobunao W, *et al.*, 2010, Gain of Nrf2 Function in Non-small-cell Lung Cancer Cells Confers Radioresistance. *Antioxid Redox Signal*, 13(11):1627–37. DOI: 10.1089/ars.2010.3219.
23. Sanchez-Vega F, Mina M, Armenia J, *et al.*, 2018, Oncogenic Signaling Pathways in the Cancer Genome Atlas. *Cell*, 173(2):321–37.
24. Singh A, Venkannagari S, Oh KH, *et al.*, 2016, Small Molecule Inhibitor of NRF2 Selectively Intervenes Therapeutic Resistance in KEAP1-Deficient NSCLC Tumors. *ACS Chem Biol*, 11(11):3214–25. DOI: 10.1021/acscchembio.6b00651.