



# Mutational Pattern in Multiple Pulmonary Nodules Are Associated With Early Stage Lung Adenocarcinoma

## OPEN ACCESS

### Edited by:

Haichang Li,  
The Ohio State University,  
United States

### Reviewed by:

Feng Wang,  
Affiliated Hospital of Nantong  
University, China  
Hsian-Rong Tseng,  
University of California, Los Angeles,  
United States

### \*Correspondence:

Chang Zou  
zou.chang@szhospital.com  
Guang-suo Wang  
wgsy01@163.com  
Chen Qiu  
szchester@163.com

†These authors have contributed  
equally to this work

### Specialty section:

This article was submitted to  
Pharmacology of Anti-Cancer Drugs,  
a section of the journal  
Frontiers in Oncology

Received: 07 September 2020

Accepted: 08 December 2020

Published: 19 February 2021

### Citation:

Dong S-w, Li R, Cheng Z, Liu D-c,  
Xia J, Xu J, Li S, Wang J, Yue Y, Fan Y,  
Cao Y, Dai L, Wang J, Zhao P,  
Wang X, Xiao Z, Qiu C, Wang G-s and  
Zou C (2021) Mutational Pattern in  
Multiple Pulmonary Nodules Are  
Associated With Early Stage Lung  
Adenocarcinoma.  
Front. Oncol. 10:571521.  
doi: 10.3389/fonc.2020.571521

Shao-wei Dong<sup>1,2†</sup>, Rong Li<sup>3†</sup>, Zhiqiang Cheng<sup>4†</sup>, Dong-cheng Liu<sup>1</sup>, Jinquan Xia<sup>1</sup>,  
Jing Xu<sup>4</sup>, Shixuan Li<sup>5</sup>, Jian Wang<sup>5</sup>, Yongjian Yue<sup>6</sup>, Yingrui Fan<sup>3</sup>, Yundi Cao<sup>3</sup>,  
Lingyun Dai<sup>7</sup>, Jigang Wang<sup>7</sup>, Pan Zhao<sup>1</sup>, Xin Wang<sup>8</sup>, Zhangang Xiao<sup>9</sup>, Chen Qiu<sup>6\*</sup>,  
Guang-suo Wang<sup>5\*</sup> and Chang Zou<sup>1,2\*</sup>

<sup>1</sup> Clinical Medical Research Centre, The First Affiliated Hospital, Southern University of Science and Technology, Shenzhen People's Hospital, Shenzhen, China, <sup>2</sup> Shenzhen Public Service Platform on Tumor Precision Medicine and Molecular Diagnosis, Shenzhen, China, <sup>3</sup> Department of Oncology, Taikang Xianlin Drum Tower Hospital, Nanjing University School of Medicine, Nanjing, China, <sup>4</sup> Department of Pathology, The First Affiliated Hospital, Southern University of Science and Technology, Shenzhen People's Hospital, Shenzhen, China, <sup>5</sup> Department of Thoracic Surgery, The First Affiliated Hospital, Southern University of Science and Technology, Shenzhen People's Hospital, Shenzhen, China, <sup>6</sup> Department of Respiratory and Critical Medicine, Shenzhen People's Hospital, Second Clinical Medical College of Jinan University, Shenzhen, China, <sup>7</sup> Department of Geriatrics, The First Affiliated Hospital, Southern University of Science and Technology, Shenzhen People's Hospital, Shenzhen, China, <sup>8</sup> Department of Biomedical Sciences, City University of Hong Kong, Hong Kong, Hong Kong, <sup>9</sup> Key Laboratory of Medical Electrophysiology of Ministry of Education, School of Pharmacy, Southwestern Medical University, Luzhou, China

The clinical significance of mutation in multiple pulmonary nodules is largely limited by single gene mutation-directed analysis and lack of validation of gene expression profiles. New analytic strategy is urgently needed for comprehensive understanding of genomic data in multiple pulmonary nodules. In this study, we performed whole exome sequencing in 16 multiple lung nodules and 5 adjacent normal tissues from 4 patients with multiple pulmonary nodules and decoded the mutation information from a perspective of cellular functions and signaling pathways. Mutated genes as well as mutation patterns shared in more than two lesions were identified and characterized. We found that the number of mutations or mutated genes and the extent of protein structural changes caused by different mutations is positively correlated with the degree of malignancy. Moreover, the mutated genes in the nodules are associated with the molecular functions or signaling pathways related to cell proliferation and survival. We showed a developing pattern of quantity (the number of mutations/mutated genes) and quality (the extent of protein structural changes) in multiple pulmonary nodules. The mutation and mutated genes in multiple pulmonary nodules are associated with cell proliferation and survival related signaling pathways. This study provides a new perspective for comprehension of genomic mutational data and might shed new light on deciphering molecular evolution of early stage lung adenocarcinoma.

**Keywords:** pulmonary nodules, whole-exome sequencing, mutation, functional analysis, pathway analysis

## INTRODUCTION

Lung adenocarcinoma (LUAD) is the most frequent subtype of lung cancer. Even with regularly performed and thorough radiological and cytological screening in high risk populations, LUAD is usually diagnosed in its middle and late stage (1). Novel therapeutic strategies have been under active research and development to conquer LUAD, but the survival rate of patients has unfortunately kept at a low level (1). The emerging of improved cancer screening methods such as liquid biopsy holds great promise for reversing this serious situation since most of solid tumors are curable in the early stage (2). Unfortunately, this strategy is still limited by the incomplete understandings of the complicated molecular mechanisms of cancer origination and malignant progression as well as intratumoral heterogeneity.

Multiple pulmonary nodules (MPNs) contribute to the majority of newly diagnosed LUAD (3). Pathologically, the malignant evolution of early stage LUAD comprises four contiguous steps: atypical adenomatous hyperplasia (AAH), *in situ* pulmonary adenocarcinoma (AIS, also named as bronchioloalveolar carcinoma, BAC), minimally invasive adenocarcinoma (MIA), and invasive adenocarcinoma (IA), all of which could be well exemplified in some patients with MPNs (4). It is well characterized that AAH represents the pre-cancerous lesion of LUAD and AIS is the localized adenocarcinoma without any invasion. MIA and IA are differentiated by the tumor size. MIA is defined as single or multiple invasive neoplasia less than 5 mm and restrained in stage of IA, while IA represents invasive tumor larger than 5 mm (5).

In view of different pathological grades and clinical stages of individual adenoma loci in the same genetic background, MPNs represent an ideal model for biological characterization of molecular evolution of the early stage LUAD. A number of efforts have been made in depicting the molecular diagram of MPNs by using genome-based approaches, e.g., whole genome sequencing, whole exome sequencing (WES) and methylation analysis (6–9). However, the theoretical and clinical significance of these reported mutations are largely limited by the lack of paired validation studies due to the rarity of malignant cells as well as the paucity of neoplasia in MPNs samples (10). In addition, since the origination and malignant progression of solid tumors is a multifaced and multiple-step orchestration, these findings, mostly based on single gene change-directed analysis, needs to be carefully interpreted and reevaluated especially for clinical application.

To better understand the molecular evolution of MPNs, we performed WES in 21 tissues dissected from 4 patients with MPNs and analyzed the somatic mutations using a cell function and pathway converging-directed strategy. Mutated genes e.g., KTM2C and EGFR, and patterns of these mutations were successfully identified and characterized. No ubiquitous mutated genes or mutation loci were discovered in all malignancy or all patients. Interestingly, more mutated genes were found in IA samples than in MIA samples. KEGG (Kyoto Encyclopedia of Genes and Genomes, [www.genome.jp](http://www.genome.jp)) and GO

(Gene Ontology, [www.geneontology.org](http://www.geneontology.org)) analysis indicates that mutated genes tend to converge into cell proliferation and survival related signaling pathways in malignancy in comparison with adjacent normal samples. This converging tendency seems to be more obvious in IA than in MIA. Moreover, the levels of protein structural changes caused by different mutations as well as that of converged signaling pathways are higher in IA than in MIA. The levels of protein structural changes and converged signaling pathways are also higher in all malignant lesions than in adjacent normal tissues, indicating these protein molecules and pathways may be responsible for the malignant evolution of LUAD. Some specific driver genes in cancer development such as EGFR, mutations of this gene are found both in cancer lesions and adjacent normal samples, despite of more severe extent of mutations in malignancy, suggesting driver gene mutations need to be further validated particularly for clinical application. Besides, in comparison with The Cancer Genome Atlas database (TCGA), most of the mutated genes in MIA and IA identified in this study are genes functionally related to the development of lung cancers. Our findings, by function-directed analysis of WES data from MPNs, that a group of functionally converged mutations rather than a single driver gene mutation might be the triggering onset of the origination and malignant progression of lung adenocarcinoma. This study will provide new insights and new strategies to decipher the early development of lung cancer.

## MATERIALS AND METHODS

### Patients and Clinical Information

Tumor samples and their matched normal tissues were obtained from four patients from Shenzhen people's hospital from 2017 to 2019. The clinical information such as age/gender/smoking history was listed in **Supplementary Table 1**. These patients did not receive preoperative chemotherapy or radiotherapy before surgery. Tumor sizes ranged from 0.2 to 3.0 cm according to pathology reports. All patients were free of extra thoracic metastasis. Three of the patients were non-smokers, and one patient had a history of cigarette smoking. The images of CT and hematoxylin-eosin staining were reviewed by experienced pathologists to determine the histopathological subtype and were shown in **Supplementary Figure S1**. The present study was approved by the Ethics Committee of Shenzhen people's hospital. All study procedures were performed according to the Declaration of Helsinki ethical principles. Informed consents were obtained from the patients.

### Sample Preparation

Genomic DNA from formalin-fixed paraffin-embedded (FFPE) samples was extracted using QIAamp DNA FFPE Tissue Kit (Qiagen, Valencia, CA). DNA from fresh tumor tissue and blood was extracted by DNAeasy Blood and Tissue kit (Qiagen, Valencia, CA). Genomic DNA was fragmented into ~250 bp by M220 Focused-ultrasonicator (Covaris, Inc. Woburn, MA)

followed by whole genome library preparation using KAPA Hyper Prep Kit (KAPA Biosystems, Wilmington, MA). Exome capture was performed using the SureSelect Human All Exon V6 panel (Agilent Technologies, Santa Clara, CA).

## Whole Exome Sequencing (WES)

WES was performed by Geneseq Technology Inc. (Nanjing, China) using an Illumina HiSeq4000 system. The adapters in FASTQ files were cleaned using Trimmomatic-0.36 (11). The alignment was performed using Burrows-Wheeler Aligner (12), and GRCh37 (hg19) downloaded from Ensembl website was used as reference. The exome mutations in pulmonary nodules and para-cancer (PAR) samples were compared with mutations obtained from peripheral blood mononuclear cells (PBMCs), and only somatic specific mutations were retained for further analysis.

## Converging Analysis

Converging index is defined as follows: 1) take any given two samples S1 and S2 with mutated genes from some pathways, KEGG for example 2) the numbers of mutated pathways (pathways with mutated genes) were calculated as P1 and P2; 3) the number of overlapped mutated pathways between two samples (mutated pathways occurred in both samples) is counted as P; and 4) the converging index is calculated as  $(P/P1) \times (P/P2)$ . The converging index in GO biological pathways and GO molecular functions is calculated using the same method.

The ranking of converged pathways/functions was guided by the probability of Poisson distribution, which is calculated using the number of mutations involved in certain pathways in PAR and IA samples. The Poisson distribution is performed in R 3.6.3 using `ppois()` function. The KEGG pathway information of human proteome was downloaded from the KEGG website ([www.genome.jp](http://www.genome.jp)) and the human GO biological pathway and molecular function information were retrieved from the Uniprot website ([www.uniprot.org](http://www.uniprot.org)).

## EGFR Protein Structural Analysis

EGFR protein structure was retrieved from the Protein Data Bank ([www.rcsb.org](http://www.rcsb.org)) using entry 1M17. The protein structures and mutated residues were illustrated by Pymol software (The PyMOL Molecular Graphics System, Version 1.2r3pre, Schrödinger, LLC.) using “cartoon” and “surface” functions.

## Comparison With TCGA Mutation Data

The genomic mutation information in 32 types of cancers was retrieved from the TCGA (The Cancer Genome Atlas) (13) via Xena platform from University of California Santa Cruz. The number of mutations in samples were compared with the TCGA genomic mutation data in each types of cancer, and a paired Wilcoxon test was performed to evaluate the difference between each types of samples.

## Protein-Protein Interaction Network (PPIN) Analysis

The protein-protein interaction information of mutated genes was retrieved from BioGrid (The Biological General Repository

for Interaction Datasets) database (14), and the PPIN in each types of samples were illustrated using Cytoscape (15).

## Statistics

All the statistics are performed by RStudio ([www.rstudio.com](http://www.rstudio.com)) using R version 3.6.1, and a p-value cutoff at 0.05 is used in all the analysis.

## RESULTS

### Characterization of Pulmonary Nodule Samples

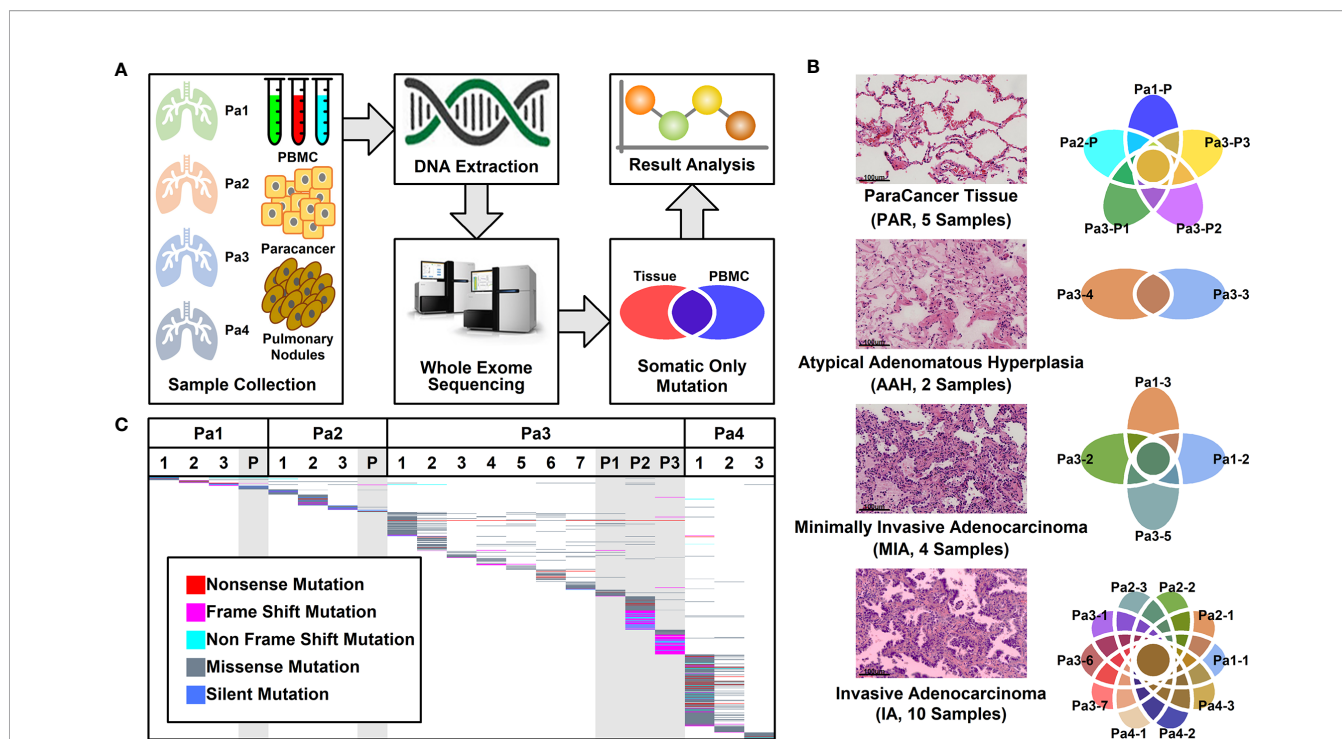
A total of 16 nodules with different stages of neoplasia and five para-cancer (PAR) tissues from four patients diagnosed as multiple pulmonary nodules (MPNs) were collected during surgery. After DNA extraction, whole exome sequencing (WES) was performed on these samples and peripheral blood mononucleated cells (PBMCs) of patients. The somatic mutations were retrieved after the filtration against germ-line mutations through a comparison with PBMC mutations, as illustrated in **Figure 1A**.

Among these 21 samples, two were in atypical adenomatous hyperplasia (AAH) stage, four in minimally invasive adenocarcinoma (MIA) stage, 10 in invasive adenocarcinoma (IA) stage, and five were PAR tissues. The histological characteristics of samples together with their IDs from each stage were presented in **Figure 1B**. The locations of all the lesions were indicated in the CT images and the histological image of each tissue were shown in **Supplemental Figure 1**.

### Mutational Analysis of Whole Exome Sequencing Results

WES was performed on these 21 samples from four patients as well as their PBMCs. The somatic mutations in each sample were summarized in **Figure 1C**. Most of the mutations are sample specific, with a small portion of mutations shared in more than one sample (scattered lines in **Figure 1C**). The number of mutations varies among different samples, with the least number of seven mutations (Pa1-1; Pa-2; Pa2-P) and the highest number of 214 mutations (Pa4-1). The average number of mutations in PAR, AAH, MIA, and IA samples are 49.4, 31.5, 29.5, and 53.0, respectively, and the average number of mutated genes in PAR, AAH, MIA, IA samples are 44.2, 25.5, 24.8, and 43.8, respectively, as listed in **Figures 2A, B**. These results indicated that there is no significant difference in the number of mutations or mutated genes between AAH and MIA samples, but there is a dramatic increase in the number of both mutations and mutated genes in IA samples as compared to the rest of samples.

For mutations on individual gene, missense mutation is the most common one across all types of samples (**Supplemental Figure 2**). As for the mutation types, single nucleotide polymorphism (SNP) is the most common type of mutations across all the samples (**Supplemental Figure 3**), which is consistent with the prevalence of missense mutations; referring



**FIGURE 1** | Pulmonary nodule sample selection and whole exome sequencing. **(A)**, Schematic illustration of research design. Twenty-one tissue samples and PBMC samples were collected. After DNA extraction and whole exome sequencing (WES), somatic mutations in different tissues were further analyzed; **(B)** Histological characteristic representation of samples and patient IDs with histopathological stages; **(C)** Summary of WES results across all samples. Each different-colored line represents a different type of mutations, and the scattered lines represent the shared mutations between different samples.

to SNV (Single Nucleotide Variant) classes in SNPs, C→T and G→A conversions are the top two common SNV types, and these two types of conversions occur most in IA samples (**Supplemental Figure 4**).

### Hotspot Mutated Gene Analysis in Pulmonary Nodules

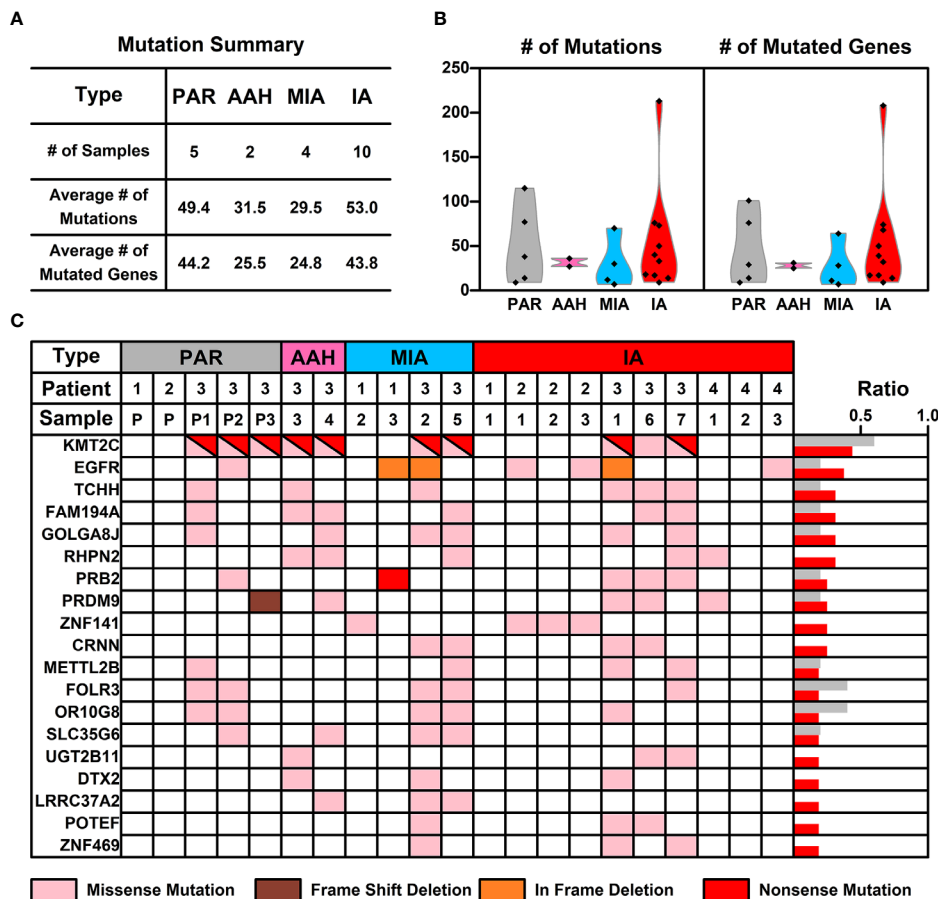
The frequency of mutated genes in all nodules (AAH, MIA, and IA samples) were further counted and ranked, and the hotspot genes (mutated genes occurred in more than 2 pulmonary nodule samples) were listed in **Figure 2C**. Among these 19 mutated hotspot genes, *KMT2C* (with both missense and nonsense mutations) gene mutation has the highest overall frequency, however, this mutation was only found in the Pa3 samples (including PAR samples); *EGFR* gene mutation, with two types of mutations (missense and in frame deletion mutations), occurs in all four patients and two types of pulmonary samples (MIA and IA); *PRB2* gene mutation, with two types of mutations (missense and nonsense mutations), occurs in two of the patients and two types of pulmonary samples (MIA and IA). Interestingly, some of these hotspot genes with mutation type changes in different types of samples were found and will be further discussed in the following sections.

An interesting phenomenon, as discovered in our study, as well as in many other studies on pulmonary nodules, is the low concurrence of mutated genes between different samples. For example, mutated *EGFR*, the gene with the highest frequency

(*KMT2C* is treated as specific in patient 3, and hence not counted here), only occurs in 37.5% of pulmonary nodules (6 out of 16 samples). The low concurrence of mutated genes indicates the high genetic diversity of pulmonary nodules, and hence we took a step back and tried to explore the commonality among pulmonary nodule samples from a perspective of signaling pathways and cell functions.

### Converging Tendency Towards Functions or Pathways Related to Malignant Transformation in Pulmonary Nodules

To explore whether there is a converging tendency of mutated genes in samples at different nodules, we introduced a converging index, which is defined as the overlapping of signaling pathways or cell functions, including KEGG pathways, GO Biological Pathways (GO\_BP), and GO Molecular Functions (GO\_MF), in which the mutated genes are involved between two samples (detailed in Materials and Methods). The higher converging index between two samples, the higher probability of two samples with same pathways or cell function got affected (pathway or cell function related genes got mutated). To compare the converging status between samples at different stages, we calculated the converging index for KEGG, GO\_BP, and GO\_MF between samples within each nodules and the results are shown in **Figure 3A** (there are only two samples in AAH stage hence only one converging index was generated).



**FIGURE 2** | The mutation status of samples from different nodules. **(A)** List of average numbers of mutations/mutated genes in samples from different stages; **(B)** Violin plot representing the mutations (left) and mutated genes (right) in different samples. Each dot represents the number of mutations (non-synonymous mutations, left) and mutated genes (right) in different samples. **(C)** Chart listing hotspot mutated genes occurring in more than two pulmonary nodules. Each colored rectangle represents a mutated gene with different colors indicating different mutation types.

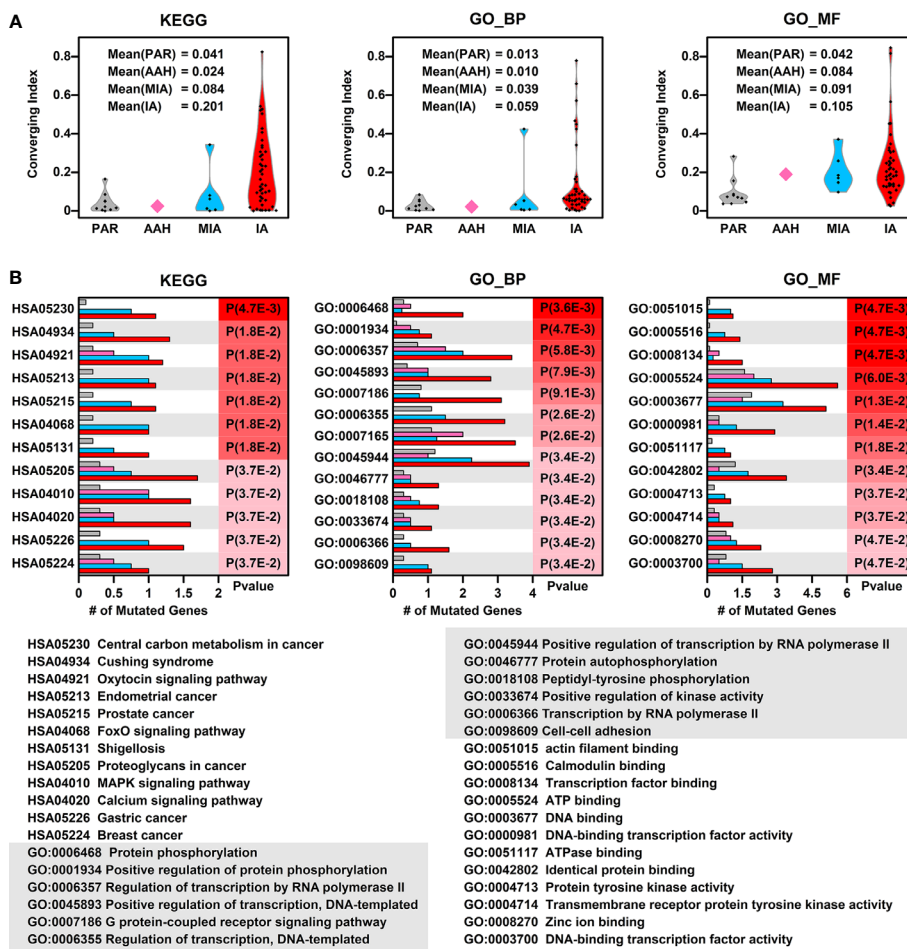
In all three types of pathway or cell function terms (KEGG, GO\_BP, and GO\_MF), the average converging indexes of lesions are higher than those of PAR tissues, except for the indexes in AAH samples, which might be caused by the limitation of sample numbers. Moreover, the indexes increase with the developmental stages of pulmonary nodules (AAH<MIA<IA).

To further explore the converged status of nodular samples at different stages, we retrieved all converged functions/pathways, ranked them based on the enrichment levels (detailed in Materials and Methods) (Figure 3B). Among the top 12 converged KEGG pathways, most of them are cancer related or important signaling pathways such as FoxO/MAPK/Calcium pathways; among the top 12 converged GO biological pathways, most of them are involved in transcription regulation and protein phosphorylation, the aberrant functions of these pathways have been shown directly related to lung cancer development in many studies (16). Moreover, GO:0098609 Cell-cell adhesion, which is directly related to epithelial-mesenchymal transition (EMT), is also converged in

lesions, indicating that the cell-cell adhesion ability has been affected in the pathogenesis of LUAD and might contribute to metastasis. Among the top 12 converged GO molecular functions, many of them are related to regulation of transcription, which is in consistent with GO\_BP results. In addition, GO:0004713, Protein tyrosine kinase activity and GO:0004714 Transmembrane receptor protein tyrosine kinase activity, that are directly involved in cancer-related signaling, have also been affected in pulmonary neoplasia. These findings implicate that multiple cell function disorders in these early stage lesions might trigger and/or accelerate the development of LUAD.

### Effects of Mutation on the Function of the Proteins Encoded by Hotspot Genes

Aside from converging tendency of mutated genes in cancer-related cellular functions and pathways, we also investigated the effect of mutation on the function of hotspot genes at the protein level. The changes of mutation types in pulmonary nodules were



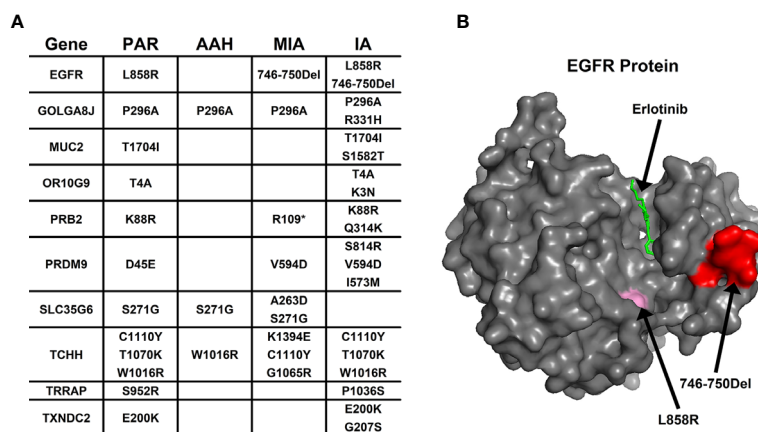
**FIGURE 3** | Converging status of samples from different nodules. **(A)** The converging ratio of KEGG pathways (left), GO\_BP (GO biological pathways, middle), and GO\_MF (GO molecular functions, right) in samples from different nodules. The converging ratio is defined in Materials and Methods with higher converging index representing the higher tendency of mutated genes converging into the same pathway or function; **(B)** The converged KEGG pathways (left), GO\_BP (middle), and GO\_MF (right) with the number of mutated genes in samples from different stages (PAR, grey; AAH, pink; MIA, cyan; IA, red) and their names (bottom). The converged items were ranked based on the probability calculated by Poisson distribution using the number of mutated genes in PAR samples and IA samples.

compared with mutations in PAR tissues listed in **Figure 4A**. For example, in EGFR protein, there is a point mutation (L858R) in PAR tissues and IA samples; another mutation type, a deletion of 5 amino acids (746-750Del), occurs in both MIA and IA samples. 746-750Del locates in Erlotinib binding region (highlighted in red, **Figure 4B**), and the deletion of this beta3-alphaC loop region has been previously reported in Erlotinib-resistant patient (17).

In PRB2 protein, there is a point mutation of K88R in PAR tissues and IA samples; a nonsense mutation (R1098\*) occur in MIA and IA samples. The nonsense mutation R1098\* is a premature stop codon which might affect the function of PRB2 protein, and might relate to the development of pulmonary nodules.

Missense mutation might also cause dramatic structural changes. For example, in OR10G9 protein, a missense mutation (K3N) occurs in IA samples, leading to conversion of a positively charged amino acid (Lysine) to a polar amino acid

(Asparagine). In PRB2 protein, a missense mutation (Q314K) occurs in IA samples, leading to conversion of a polar amino acid (Glutamine) to a positively charged amino acid (Lysine). In PRDM9 protein, two missense mutations occur in MIA and IA samples (V594D; S814R) leading to conversion of a non-polar amino acid (Valine) to a negatively charged amino acid (Aspartic acid) and a polar amino acid (Serine) to a positively charged amino acid (Arginine). In SLC35G6 protein, a missense mutation occurs in MIA samples (A263D), leading to conversion of a non-polar amino acid (Alanine) to a negatively charged amino acid (Aspartic Acid). In TCHH protein, a missense mutation occurs in MIA samples, (K1394E) leading to conversion of a positively charged amino acid to a negatively charged amino acid (Glutamic Acid). In TXNDC2 protein, a new missense mutation (G207S) occurs in IA samples, leading to conversion of a nonpolar amino acid (Glycine) to a polar amino acid (Serine).



**FIGURE 4** | The mutation change in the same gene from different pulmonary nodules. **(A)** List of genes with different mutation in para-cancer tissues (PAR) and pulmonary nodule tissues (AAH, MIA and IA); **(B)** The 3D surface structure of EGFR in complex with Erlotinib. The different types of mutations are highlighted in pink and red colors.

## Pulmonary Nodule Mutated Genes

To explore the association of cancer related genes and mutated genes in different pulmonary nodules screened in WES, we retrieved the mutational data of these genes in 32 types of cancers from TCGA database, and compared the average number of mutations per gene among different types of cancers. Among 32 types of cancers, IA-related mutated genes have the highest number of mutations per gene in UCEC (Uterine corpus endometrial carcinoma), LUSC (Lung squamous cell carcinoma) and LUAD (Lung adenocarcinoma) ranks among the top #5 and #7, respectively of these 32 types of cancers, indicating the IA-related mutated genes have a higher mutation rate in lung cancers comparing to other types of cancers (**Figure 5A**). Among the four stages of nodules, pulmonary nodule related genes have a higher mutation rate in all types of cancers in comparison with PAR-related genes, and the mutation rate increases with the development of pulmonary nodules (IA>MIA, IA>AAH, AAH>PAR) (**Figure 5B**).

To explore the functional importance of pulmonary nodule related mutated genes, we calculated the number of each mutated gene involved in KEGG pathways (**Figure 5C**), GO\_BP (**Figure 5D**), and GO\_MF (**Figure 5E**). The rationale behind this is that the higher the number of genes involved in the pathways/functions, the more important role one gene could play, and hence the mutation could make bigger impact on cell activities. In general, the average numbers of mutated genes involved in pathways/functions are higher in pulmonary nodules than that of PAR-related mutated genes, suggesting the increased functional importance of mutated genes in pulmonary nodules.

We further constructed the protein-protein interaction (PPI) networks generated by mutated genes from different pulmonary nodules (**Figure 5F**) and compared them with each other (the PPI network of AAH-related genes was not shown due to limited number of mutated genes). The PPI network generated by IA-related mutated genes is more complex when compared with MIA or PAR nodules, indicating a complex function affected in

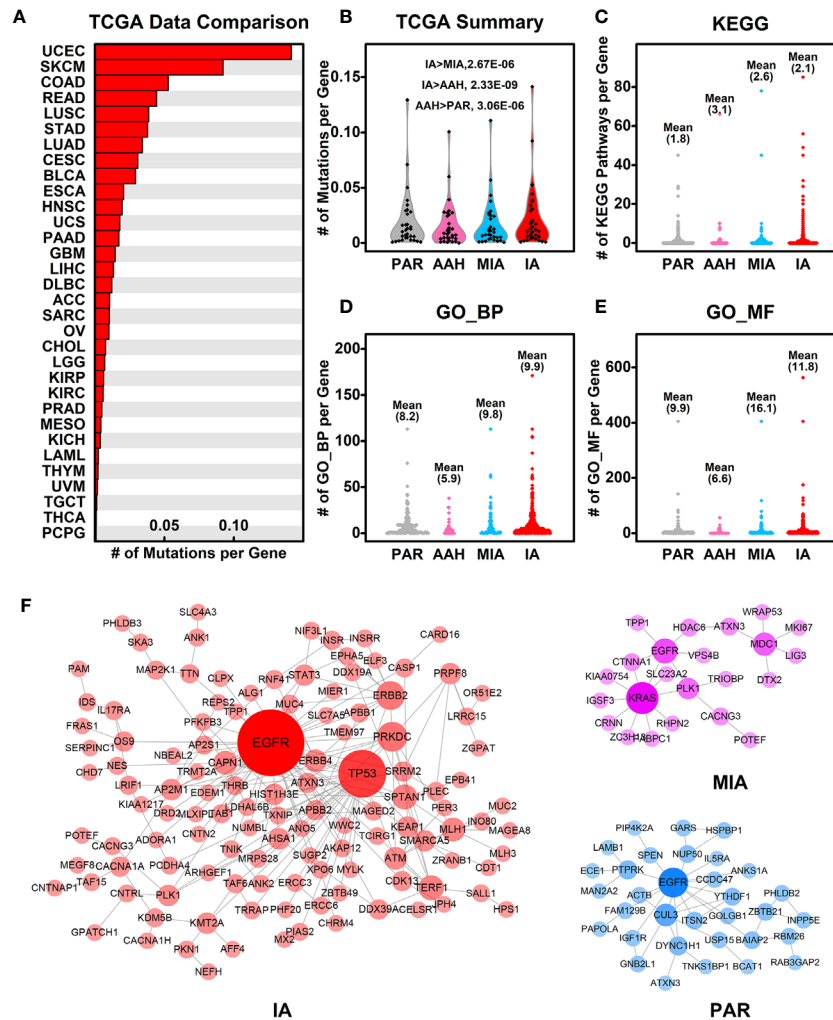
IA samples. The complexity of PPI network could contribute to the explanation of development of pulmonary nodules.

## DISCUSSION

MPNs is defined as more than one solid nodules in the lung of a single patient. Pathologically, MPNs are commonly identified as AAH, AIS, MIA, and IA. They represent an ideal model of deciphering the clonal evolution of early stage lung adenocarcinoma. Understanding the gene mutations in MPNs holds great promise for prevention and treatment strategy of lung adenocarcinoma.

Accumulating mutations as well as genes were identified by using next generation sequencing on MPNs samples (6, 9, 18). However, these studies mainly used single gene change-directed data analytic strategy and the clinical significance of these findings are largely limited (19). To overcome the limitations of current genome-based data, we analyzed the somatic mutations identified in the WES data from 21 MPNs followed by a cell function and pathway converging-directed analysis. Based on this function-directed analysis, we found that the extent of protein structural changes caused by different mutations as well as the amounts of affected pathways/functions in all lesions are positively correlated with the degree of malignancy, indicating that during the malignant evolution, comprehensive mutations in several key signaling pathways other than single gene are accumulated, which might endow certain cell populations more aggressive phenotype in terms of accelerated proliferation rates and the malignant progression of lesions.

Among the mutated genes identified in all samples, KMT2C other than EGFR was found in 10 out of 21 lesions. This result is in consistent with the recently reported finding from cell line experiments, indicating that KTM2C might be a novel driver gene correlated with KRAS mutation during the development of early stage lung adenocarcinoma (20). However, KMT2C mutation is only found in one of four patient and only one IA out of all 10



**FIGURE 5 |** Association of mutated genes in pulmonary nodules and other cancer genes. **(A)** The number of mutations per gene in different types of cancers. The mutated genes in IA samples were used here, and the mutation data from different types of cancers were retrieved from TCGA database; **(B)** The summary of mutations per gene in different pulmonary nodules compared with different types of cancers. With the development of pulmonary nodules, the chances of pulmonary nodule related genes that are mutated in different types of cancers increased; **(C–E)** The number of KEGG pathways **(C)**, GO\_BP **(D)**, and GO\_MF **(E)** in which each mutated gene involved is plotted in different nodules. With the development of pulmonary nodules, the average number of mutated gene involved pathways/functions increases, indicating the functional importance of mutated genes; **(F)** The PPI networks generated using mutated genes from PAR, MIA, and IA samples.

samples harbors a missense mutation, all other samples in particular adjacent normal samples harbored with nonsense mutation. Moreover, many of these mutations including those in EGFR were also identified in adjacent normal tissues when germ-line mutations were filtered during the process of analysis. These findings indicate the complicated background of cancer lesions and the clinical significance of single gene mutation in samples except of cancerous tissues should be carefully reevaluated. On the contrary, EGFR mutation are mainly found in malignant lesions, among which in frame deletion is the major mutational pattern. Similar results were obtained in the amounts and patterns of many other gene mutations in this study.

The KEGG signaling pathway analysis indicates that varied mutated genes tend to converge into signaling pathways related

to pivotal cell growth regulation in malignancy in comparison with adjacent normal samples. This converging trend seems to be more obvious in IA than in MIA, indicating a cancer-like microenvironment forced selection of functional mutations during the origination and development of early stage lung adenocarcinoma. Considering the similar number of mutations/mutated genes in PAR and IA samples, the increased converging index indicating mutated genes tend to focus on certain functions/pathways, or only cells with certain impaired functions can survive under these microenvironments. Among the top 12 converged KEGG pathways, most of them are cancer related or important signaling pathways such as FoxO/MAPK/Calcium pathways. Among the top 12 converged GO biological pathways, most of



them are involved in transcription regulation and protein phosphorylation, the aberrant functions of these pathways have been shown directly related to cancer development in many studies. Moreover, the levels of protein structural changes caused by mutations as well as the extent of mutations in these converged pathways are higher in IA as compared with that in MIA tissues, both of which are higher in malignant lesions than in adjacent normal tissues, indicating that these pathways may be responsible for the malignant evolution of pulmonary nodules.

Although whole transcriptional sequencing data from other patients might be available, lacking in-patient paired validation in protein expression levels is the major limitation of exploring WES data in MPNs studies. To circumvent this problem, we used a protein structure predicting strategy to evaluate the functional significance of hot spot gene mutations that are related to carcinogenesis and target therapy resistance (as listed in **Figure 4A**) (21). As shown in our results, 746-750 Del mutation of EGFR was identified, which might induce erlotinib resistance in NSCLC patients (17). PRB2, a tumor suppressor (22, 23), a premature stop codon was identified in our study, suggesting that its down-regulation and dysfunction might occur in patients. The mutation of TCHH is found to be related to platinum resistance in gastric carcinoma (24). In our predicting results, the conversion from positively charged amino acid to negatively charged amino acid (K1394E) of TCHH induced by missense mutation might affect its protein stability as well as the prognosis of this patient (25). Down regulation of TXNDC2, an antioxidant enzyme is directly related to cisplatin resistance in many cancer cells (26, 27). In our findings, conversion from nonpolar amino acid to polar one (G207S) in protein structure is also predicted. These findings are coincide with previous evidences on the mutation analysis in a serial of studies (28–31). Besides, in comparison with The Cancer Genome Atlas database (TCGA), most of mutated genes in MIA and IA identified in this study are genes functionally related to the development of lung cancers.

In this study, the findings by function-directed analysis of WES data from MPNs suggested that a group of functionally converged mutations rather than a single driver gene mutation might be the trigger of the origination and malignant progression of lung adenocarcinoma. Our results indicated that other than well-known genes such as EGFR, KRAS, and TP53, mutations in other genes involved in relating pathways (as listed in **Figure 3B**) deserve more attention, and an involvement of these genes might be necessary in the future screening of early stage lung adenocarcinoma. Besides, our results shed some new insights as well as new strategy to decipher the early development of lung cancer, and also provide many candidate genes which might be used as potential drug targets in the future treatment of lung adenocarcinoma.

## DATA AVAILABILITY STATEMENT

The data presented in the study are deposited in Genome Sequence Archive for Human (“<https://bigd.big.ac.cn/gsa-human/>”) repository, accession number HRA000465.

## ETHICS STATEMENT

The present study was approved by the Ethics Committee of Shenzhen People’s Hospital. The patients/participants provided their written informed consent to participate in this study. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

## AUTHOR CONTRIBUTIONS

CZ, S-WD, CQ, and G-sW conceived the research idea. CZ, S-WD, ZC, and CQ prepared and wrote the manuscript. D-CL, JXu, SL, PZ, and JAW collected the clinical samples. RL, JXi, YY, JGW, LD, and XW analyzed the WES data. SD, YF, YC, and CZ revised the manuscript. All authors contributed to the article and approved the submitted version.

## FUNDING

This study is funded by The Science and Technology Foundation of Shenzhen (JCYJ20180305164128430); the International Cooperation Foundation of Shenzhen (GJHZ20180928171602104); the Shenzhen Economic and Information Committee “Innovation Chain and Industry Chain” integration special support plan project (20180225112449943); the Shenzhen Public Service Platform on Tumor Precision Medicine and Molecular Diagnosis; Research Grants Council of the Hong Kong Special Administrative Region, China (Project No. CityU 21101115, 11102317, 11103718, 11103619, R4017-18, C4041-17GF), a grant from Guangdong Basic and Applied Basic Research Foundation (Project No. 2019B030302012), Young Scientists Fund of the National Natural Science Foundation of China (81802384); International innovation team for early diagnosis and precise treatment of lung cancer (2016, KQTD2016113015442590); Scientific Research Project of Taikang Xianlin Drum Tower Hospital (TKKY20193805).

## ACKNOWLEDGMENTS

We would like to thank the staffs of Department of Pathology in Shenzhen People’s Hospital for technical support and the nurses of Department of Thoracic Surgery in Shenzhen People’s Hospital for taking care of patients.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fonc.2020.571521/full#supplementary-material>

## REFERENCES

- Siegel RL, Miller KD, Jemal A. Cancer statistics. *CA Cancer J Clin* (2016) 69(1):7–34. doi: 10.3322/caac.21551
- Santarpia M, Liguori A, D'Aveni A, Karachaliou N, Gonzalez-Cao M, Daffinà MG, et al. Liquid biopsy for lung cancer early detection. *J Thorac Dis* (2018) 10(3):S882–97. doi: 10.21037/jtd.2018.03.81
- Feng RM, Zong YN, Cao SM, Xu RH. Current cancer situation in China: Good or bad news from the 2018 Global Cancer Statistics? *Cancer Commun* (2019) 39(1):1–12. doi: 10.1186/s40880-019-0368-6
- Zheng M. Classification and Pathology of Lung Cancer. *Surg Oncol Clin N Am* (2016) 25(3):447–68. doi: 10.1016/j.soc.2016.02.003
- He P, Yao G, Guan Y, Lin Y, He J. Diagnosis of lung adenocarcinoma in situ and minimally invasive adenocarcinoma from intraoperative frozen sections: an analysis of 136 cases. *J Clin Pathol* (2016) 69(12):1076–80. doi: 10.1136/jclinpath-2016-203619
- Hu X, Fujimoto J, Ying L, Fukuoka J, Ashizawa K, Sun W, et al. Multi-region exome sequencing reveals genomic evolution from preneoplasia to lung adenocarcinoma. *Nat Commun* (2019) 10(1):2978. doi: 10.1038/s41467-019-10877-8
- Li Y, Li X, Li H, Zhao Y, Liu Z, Sun K, et al. Genomic characterisation of pulmonary subsolid nodules: mutational landscape and radiological features. *Eur Respir J* (2020) 55(2):1901409. doi: 10.1183/13993003.01409-2019
- Ren Y, Huang S, Dai C, Xie D, Zheng L, Xie H, et al. Germline Predisposition and Copy Number Alteration in Pre-stage Lung Adenocarcinomas Presenting as Ground-Glass Nodules. *Front Oncol* (2019) 9:288. doi: 10.3389/fonc.2019.00288
- Taylor TD, Rao X, Campa MJ, Wang J, Gregory SG, Patz EFJ. Whole Exome Sequencing of Cell-Free DNA for Early Lung Cancer: A Pilot Study to Differentiate Benign From Malignant CT-Detected Pulmonary Lesions. *Front Oncol* (2019) 9:317. doi: 10.3389/fonc.2019.00317
- Izumchenko E, Chang X, Brait M, Fertig E, Kagohara LT, Bedi A, et al. Targeted sequencing reveals clonal genetic changes in the progression of early lung neoplasms and paired circulating DNA. *Nat Commun* (2015) 6:1–13. doi: 10.1038/ncomms9258
- Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* (2014) 30(15):2114–20. doi: 10.1093/bioinformatics/btu170
- Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* (2009) 25(14):1754–60. doi: 10.1093/bioinformatics/btp324
- Weinstein JN, Collisson EA, Mills GB, Shaw KRM, Ozenberger BA, Ellrott K, et al. The Cancer Genome Atlas Pan-Cancer analysis project. *Nat Genet* (2013) 45(10):1113–20. doi: 10.1038/ng.2764
- Oughtred R, Stark C, Breitkreutz BJ, Rust J, Boucher L, Chang C, et al. The BioGRID interaction database: 2019 update. *Nucleic Acids Res* (2019) 47(D1):D529–41. doi: 10.1093/nar/gky1079
- Shannon P, Markiel A, Owen O2, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* (2003) 13:2498–504. doi: 10.1101/gr.1239303
- Hanahan D, Weinberg RA. Hallmarks of cancer: The next generation. *Cell* (2011) 144(5):646–74. doi: 10.1016/j.cell.2011.02.013
- Santoni-Rugiu E, Grauslund M, Melchior LC, Costa JC, Sørensen JB, Urbanska EM. Heterogeneous resistance mechanisms in an EGFR exon 19-mutated non-small cell lung cancer patient treated with erlotinib: Persistent FGFR3-mutation, localized transformation to EGFR-mutated SCLC, and acquired T790M EGFR-mutation. *Lung Cancer* (2017) 113:14–7. doi: 10.1016/j.lungcan.2017.08.024
- Liang W, Zhao Y, Huang W, Gao Y, Xu W, Tao J, et al. Non-invasive diagnosis of early-stage lung cancer using high-throughput targeted DNA methylation sequencing of circulating tumor DNA (ctDNA). *Theranostics* (2019) 9(7):2056–70. doi: 10.7150/thno.28119
- Qiu T, Li W, Zhang F, Wang B, Ying J. Major challenges in accurate mutation detection of multifocal lung adenocarcinoma by next-generation sequencing. *Cancer Biol Ther* (2020) 21(2):170–7. doi: 10.1080/15384047.2019.1674070
- Fang B. RAS signaling and anti-RAS therapy: lessons learned from genetically engineered mouse models, human cancer cells, and patient-related studies. *Acta Biochim Biophys Sin (Shanghai)* (2016) 48(1):27–38. doi: 10.1093/abbs/gmv090
- Sigismund S, Avanzato D, Lanzetti L. Emerging functions of the EGFR in cancer. *Mol Oncol* (2018) 12(1):3–20. doi: 10.1002/1878-0261
- Caputi M, Groeger AM, Esposito V, De Luca A, Masciullo V, Mancini A, et al. Loss of pRb2/p130 expression is associated with unfavorable clinical outcome in lung cancer. *Clin Cancer Res* (2002) 8(12):3850–6.
- Cito L, Indovina P, Forte IM, Pentimalli F, Di Marzo D, Somma P, et al. pRb2/p130 localizes to the cytoplasm in diffuse gastric cancer. *J Cell Physiol* (2015) 230(4):802–5. doi: 10.1002/jcp.24805
- Yang Q, Zhu C, Zhang Y, Wang Y, Wang Y, Zhu L, et al. Molecular analysis of gastric cancer identifies genomic markers of drug sensitivity in Asian gastric cancer. *J Cancer* (2018) 9(16):2973–80. doi: 10.7150/jca.25506
- Kuo L-H, Li J-H, Kuo H-T, Hung C-Y, Tsai H-Y, Chiu W-C, et al. Effect of charged amino acid side chain length at non-hydrogen bonded strand positions on  $\beta$ -hairpin stability. *Biochemistry* (2013) 52(44):7785–97. doi: 10.1021/bi400911p
- Pereira L, Igea A, Canovas B, Dolado I, Nebreda AR. Inhibition of p38 MAPK sensitizes tumour cells to cisplatin-induced apoptosis mediated by reactive oxygen species and JNK. *EMBO Mol Med* (2013) 5(11):1759–74. doi: 10.1002/emmm.201302732
- Qiao B, Jiménez-Ángeles F, Nguyen TD, De La Cruz MO. Water follows polar and nonpolar protein surface domains. *Proc Natl Acad Sci U S A* (2019) 116(39):19274–81. doi: 10.1073/pnas.1910225116
- Betje J, Schneider NI, Harbaum L, Pollheimer MJ, Lindtner RA, Kornprat P, et al. MUC1, MUC2, MUC5AC, and MUC6 in colorectal cancer: expression profiles and clinical significance. *Virchows Arch* (2016) 469(3):255–65. doi: 10.1007/s00428-016-1970-5
- Houle AA, Gibling H, Lamaze FC, Edgington HA, Soave D, Fave MJ, et al. Aberrant PRDM9 expression impacts the pan-cancer genomic landscape. *Genome Res* (2018) 28(11):1611–20. doi: 10.1101/gr.231696.117
- Jethwa A, Slabicki M, Hüllein J, Jentszsch M, Dalal V, Rabe S, et al. TRRAP is essential for regulating the accumulation of mutant and wild-type p53 in lymphoma. *Blood* (2018) 131(25):2789–802. doi: 10.1182/blood-2017-09-806679
- Martincorena I, Raine KM, Gerstung M, Dawson KJ, Haase K, Van Loo P, et al. Universal Patterns of Selection in Cancer and Somatic Tissues. *Cell* (2017) 171(5):1029–41. doi: 10.1016/j.cell.2017.09.042

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Dong, Li, Cheng, Liu, Xia, Xu, Li, Wang, Yue, Fan, Cao, Dai, Wang, Zhao, Wang, Xiao, Qiu, Wang and Zou. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.