



# Genetic and microenvironmental differences in non-smoking lung adenocarcinoma patients compared with smoking patients

Qihai Sui<sup>1,2#</sup>, Jiaqi Liang<sup>1#</sup>, Zhengyang Hu<sup>1#</sup>, Zhencong Chen<sup>1</sup>, Guoshu Bi<sup>1</sup>, Yiwei Huang<sup>1</sup>, Ming Li<sup>1</sup>, Cheng Zhan<sup>1</sup>, Zongwu Lin<sup>1</sup>, Qun Wang<sup>1</sup>

<sup>1</sup>Department of Thoracic Surgery, Zhongshan Hospital, Fudan University, Shanghai, China; <sup>2</sup>Eight-Year Program Clinical Medicine, Grade of 2016, Shanghai Medical College, Fudan University, Shanghai, China

**Contributions:** (I) Conception and design: Q Sui, C Zhan; (II) Administrative support: C Zhan, Q Wang; (III) Provision of study materials or patients: J Liang, M Li; (IV) Collection and assembly of data: Z Hu, Y Huang; (V) Data analysis and interpretation: Z Chen, G Bi; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

<sup>#</sup>These authors contributed equally to this work.

**Correspondence to:** Dr. Cheng Zhan; Dr. Zongwu Lin. 180 Fenglin Road, Xuhui District, Shanghai 200032, China. Email: czhan10@fudan.edu.cn; lin.zongwu@zs-hospital.sh.cn.

**Background:** Non-smoking-related lung adenocarcinoma (LUAD) has its own characteristics. Genetic and microenvironmental differences in smoking and non-smoking LUAD patients were analyzed to elucidate the oncogenesis of non-smoking-related LUAD, which will improve our understanding of the underlying molecular mechanism and be of clinical use in the future.

**Methods:** The Cancer Genome Atlas (TCGA), Gene Expression Omnibus (GEO) databases were used for clinical and genomic information. Various bioinformatics tools were used to analyze differences in somatic mutations, RNA and microRNA (miRNA) expression, immune infiltration, and stemness indices. GO, KEGG, and GSEA analyses were performed with R. A merged protein-protein interaction (PPI) network was constructed and analyzed. A miRNA-differentially expressed gene network was constructed with miRNet. qRT-PCR was used for validation of 4 most significantly differently expressed genes and 2 miRNAs in tumor samples obtained from 20 pairs of non-smoking and smoking patients.

**Results:** Five hundred and one patients with LUAD were obtained, including 210 in the non-smoking group and 292 in the smoking group. A total of 174 significantly altered somatic mutations were detected, including mutations in tumor protein p53 and epidermal growth factor receptor, which were downregulated in non-smoking-related LUAD. At the RNA level, 231 significantly differentially expressed genes were obtained; 124 were upregulated and 107 downregulated in the non-smoking group. GSEA analysis revealed 42 significant pathways. Other functional and enrichment analyses of somatic mutations and RNA expression levels revealed that these genes were significantly enriched in receptor activity regulation and receptor binding. Differences in microenvironments including immune infiltration (e.g., CD8<sup>+</sup> T cells and resting mast cells) and stemness indices were also found between groups. A 79-pair interaction was found between differentially expressed genes and miRNAs, of which miR-335-5p and miR-34a-5p were located in the center. Twenty-one genes, including vitronectin, neurotensin, and neuronatin, were differentially expressed in both non-smoking LUAD patients and DMSO-treated A549 cells. And the different expression of neurotensin, neuronatin, trefoil factor family2, regenerating family member 4, miR-377-5p, miR-34a were verified with the same tendency in our own samples.

**Conclusions:** Non-smoking LUAD patients, compared to smokers, have different characteristics in terms of somatic mutation, gene, and miRNA expression and the microenvironment, indicating a diverse mechanism of oncogenesis.

**Keywords:** Lung adenocarcinoma; non-smoking; genome; microenvironment

Submitted Feb 18, 2020. Accepted for publication Jul 13, 2020.

doi: 10.21037/tlcr-20-276

View this article at: <http://dx.doi.org/10.21037/tlcr-20-276>

## Introduction

Lung cancer is the most common cancer worldwide. Non-small cell lung cancer (NSCLC) (1), the main subtypes of which include adenocarcinoma and squamous cell carcinoma, accounts for approximately 85% of the cases of lung cancer (2). Lung cancer is the most common malignant tumor and has the highest morbidity and mortality (1). According to the World Health Organization Report on the Global Tobacco Epidemic (3), there are 1.4 billion tobacco users aged 15 years and older worldwide. It is well accepted that tobacco smoking is the primary cause of lung cancer, especially lung squamous cell carcinoma (3-5) since Richard Doll and Austin Bradford Hill first proved the link between smoking and lung cancer in the 1950s (6).

However, approximately 25% of lung cancer cases worldwide, mainly adenocarcinoma, cannot be attributed to tobacco smoking; lung cancer in never smokers is the seventh leading cause of cancer deaths worldwide (7). According to clinical experience, different epidemiology and natural history are observed between lung cancers in never smokers and those in smokers (8), suggesting that lung cancer in never smokers is a 'different' disease, with specific etiology and molecular differences (9). With the advent of RNA sequencing technology, more studies have focused on specific genetic or molecular mechanisms (10); however, these have remained unclear until now in non-smoker-related lung adenocarcinoma (LUAD).

To study the differences in non-smoking compared with smoking LUAD patients, clinical data, gene expression, somatic mutations, immune infiltration, and stemness indices were analyzed to determine the role of smoking in LUAD. The effects were also confirmed in LUAD cells. This study will improve our understanding of the causes of oncogenesis of non-smoking LUAD and provide a reference for therapeutic decisions in clinical treatments.

We present the following article in accordance with

the MDAR reporting checklist (available at <http://dx.doi.org/10.21037/tlcr-20-276>).

## Methods

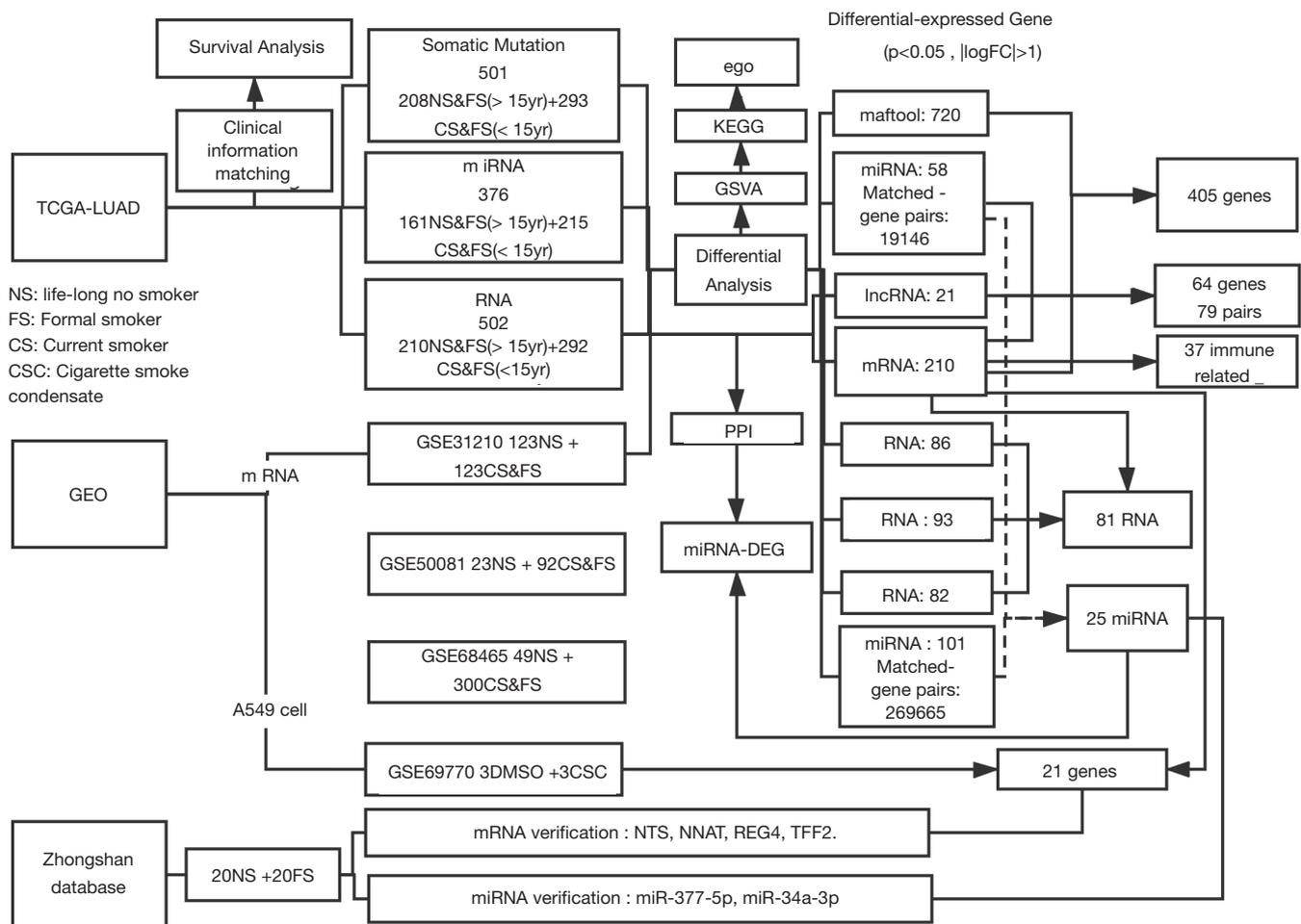
### *The Cancer Genome Atlas (TCGA) data collection*

Detailed information for patients with LUAD was downloaded from TCGA. Clinical data, RNA expression data, microRNA (miRNA) expression and somatic mutation (VarScan MAF files) data were downloaded from NCI's Genomic Data Commons (<https://gdc.cancer.gov/>) (TCGA-LUAD). After matching mRNA expression, miRNA expression, and somatic mutation data with clinical data, 502 cases of RNAs, 376 cases of miRNAs, and 501 cases of somatic mutations were obtained to explore the differences in the genomes of smoking and non-smoking LUAD patients at different levels (see *Figure 1* for details). According to the smoking data described in TCGA, patients who had never smoked or who had quit smoking for more than 15 years were classified as non-smokers while those who were current smokers, who quit smoking less than 15 years ago, or quit smoking but the years were unknown were classified as smokers (*Figure 1*).

### *GEO data collection*

The mRNA expression data of LUAD from current, former, and never smokers were download from Gene Expression Omnibus (GEO) database (<http://www.ncbi.nlm.nih.gov/geo/>). GSE31210, GSE50081, and GSE68465 datasets (11-13), published on February 20, 2008, were collected to determine the coding RNA expression signature of cigarette smoking.

The GSE69770 dataset (14), published on February 7, 2018, performed RNA sequencing of dimethyl sulfoxide (DMSO)- or cigarette smoke condensate (CSC)-treated A549 human LUAD cells. Cells were treated for 48 h or



**Figure 1** Flow diagram of whole design.

2 weeks.

**Immune infiltration data**

Immune infiltration data was collected from a previous study (15) and matched with the clinical data downloaded from the TCGA database. A total of 503 samples were finally enrolled; 210 were classified as non-smokers and 293 were classified as smokers.

**Stemness indices**

Stemness indices were collected from a previous study (16)

and matched with the clinical data downloaded from the TCGA database. A total of 496 samples were finally enrolled; 207 were classified as non-smokers and 289 were classified as smokers.

**Tissue samples**

All samples were obtained from patients at the Department of Thoracic Surgery, Zhongshan Hospital, Fudan University. Tumor samples were derived from surgical resection of LUAD between January 2019 and April 2019. All samples were quickly frozen in liquid nitrogen after removal and then stored at -80 °C before used. Finally,

20 pairs of non-smoking LUAD tissue and smoking adenocarcinoma (smoking index  $\geq 20$  pack-years) tissues from 40 patients were obtained for qRT-PCR validation.

### RNA, miRNA preparation and qRT-PCR analysis

To detect the expression of NTS, NNAT, TFF2, REG4, miR-512, miR-372 in 20 pairs of samples, RT-qPCR was carried out on an ABI Prism 7500 real-time PCR system (Applied Biosystems) with proper PCR parameters.

Total RNAs were extracted by TRIzol (Beyotime, China). The first-strand cDNA was synthesized using PrimeScript<sup>TM</sup> RT reagent Kit with gDNA Eraser (Real Time Perfect) (TaKaRa, Tokyo, Japan) according to the manufacturer's instructions. Then SYBR Premix Ex Taq<sup>TM</sup> II (Tli RNaseH Plus) (TaKaRa) was used with the following PCR parameters, 1 cycle of 30 s at 95 °C, 40 cycles of 5 s at 95 °C and 34 s at 60 °C.  $\beta$ -actin was used as the reference. Primers used in this study are listed in *Table S1*.

Total miRNAs were extracted with miRcute miRNA Isolation Kit (TianGen Biotech, Beijing, China) and then reversed by miRcute Plus miRNA First-Strand cDNA Kit (TianGen). miRcute Plus miRNA qPCR Kit (SYBR Green) (TianGen) was used in the PCR system with the following PCR parameters, 1 cycle of 2 min at 94 °C, 40 cycles of 20 s at 94 °C and 34 s at 60 °C. U6 was used as the reference. All miRNA primers were obtained from TianGen.

All the samples were repeated three times.

### Statistical analysis

The distribution of patients' characteristics (e.g., sex, race, age group, primary site, differentiation grade, and chemotherapy) was summarized using counts and percentages. Statistical analysis was performed with SPSS 23.0 software (IBM). Differential analysis of RNA, miRNA, and immune infiltration data was performed with R (version 3.6.1). All data were normalized and standardized by constructing relevant expression matrices using edgeR. Differential genes [ $P < 0.05$ , false discovery rate (FDR)  $< 0.05$  for somatic mutations, miRNAs, and others] were sorted according to logFoldChange values ( $|\log_2FC| > 1$ ) to identify significantly different expression between smoking and non-smoking patients. A merged protein-protein interaction (PPI) network was constructed and analyzed with Cytoscape software (version 3.7.2). Differences in stemness indices and verification were determined by Student's *t*-test with Graphpad Prism 8. The test level

was  $\alpha = 0.05$ , and differences were considered statistically significant at  $P < 0.05$ . Survival analysis was performed with the Kaplan-Meier method with R (version 3.6.1).

All procedures performed in this study were in accordance with the Declaration of Helsinki (as revised in 2013) and was approved by the Ethics Committee of Zhongshan Hospital, Fudan University, China (B2019 – 377R). Because of the retrospective nature of the re-search, the requirement for informed consent was waived.

## Results

*Table 1* shows the basic clinical characteristics of the patients (e.g., sex, age group, location, and stage) summarized using counts and percentages. The smokers were younger than the non-smokers. Most of the cases of LUAD in non-smokers were acinar, papillary predominant, and invasive mucinous pathology subtype. Smokers had lower 3-year survival rates [smokers (S), 72.11% vs. non-smokers (NS), 77.56%], although this difference was not statistically significant (*Figure 2*).

### Differences in somatic mutations

The distribution patterns of somatic mutations (the overall pattern is shown in *Figure S1*) were investigated between the non-smoker and smoker groups. After performing enrichment analysis, we identified 174 enriched mutations with FDR  $< 0.05$  between the non-smoker and smoker groups (*Figure 3*). The somatic mutation rates of tumor protein p53 (TP53; S, 56.31% vs. NS, 35.10%); titin (TTN; S, 57.34% vs. NS, 29.33%); KRAS proto-oncogene (KRAS; S, 31.40% vs. NS, 22.60%), and ryanodine receptor 2 (RYR2; S, 45.39% vs. NS, 22.12%), were higher in smokers, while epidermal growth factor receptor (EGFR; S, 6.83% vs. NS, 19.71%) were higher in smokers. The mutation rate of human epidermal growth factor receptor-2 (HER2; S, 17.41% vs. NS, 8.17%) and phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit alpha (PIK3CA; S, 8.60% vs. NS, 0) showed no difference between smokers and nonsmokers.

Using the somatic interactions function, which performed a pair-wise Fisher's exact test to detect significant pairs of genes, we detected the relationships among the top 50 genes with different somatic mutations (*Figure S2*). Almost all of the mutated genes were tightly linked to each other, except EGFR, which was strongly excluded. Thus, these mutations might be linked to the development of

**Table 1** Clinical features of the TCGA samples

Characteristics	Nonsmokers (n=210)		Smokers (n=292)		P value
	n	%	n	%	
Age					<0.001
≤60	36	17.14	118	40.41	
61–70	69	32.86	102	34.93	
≥71	92	43.81	66	22.60	
NA	13	6.19	6	2.05	
Sex					0.469
Female	117	55.71	153	52.40	
Male	93	44.29	139	47.60	
Anatomic location					0.587
Left	84	40.00	113	38.70	
Right	122	58.10	169	57.88	
NA	4	1.90	10	3.42	
Stage					0.366
Stage I	119	56.67	153	52.40	
Stage II	46	21.90	71	24.32	
Stage III	32	15.24	48	16.44	
Stage IV	10	4.76	16	5.48	
NA	3	1.43	4	1.37	
Pathology subtype <sup>1</sup>					0.143
Acinar predominant ADC	37	17.62	32	11.00	
Colloid ADC	0	0.00	1	0.34	
Lepidic predominant ADC	6	2.86	5	1.72	
Micropapillary predom ADC	10	4.76	14	4.81	
Papillary predominant ADC	14	6.67	9	3.09	
Solid predominant ADC	20	9.52	37	12.71	
Invasive mucinous	7	3.33	3	1.03	

<sup>1</sup>, only those with specific identification of pathological subtypes were included in this column. TCGA, The Cancer Genome Atlas; ADC, adenocarcinoma.

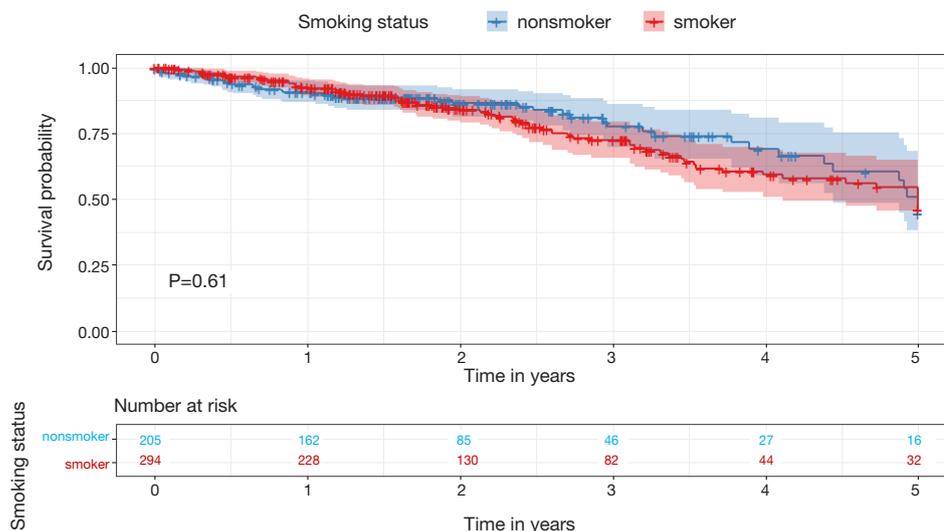
LUAD.

### ***Differentially expressed coding genes and functional annotation in TCGA-LUAD***

To determine whether smoking can change biological characteristics, expression profiles were examined for differentially expressed genes (DEGs). Of the 60,483 ensemble numbers, 13,747 were successfully converted

to gene names by org.Hs.eg.db, and the edgeR algorithm identified 231 DEGs [210 mRNAs and 21 long non-coding RNAs (lncRNAs)] ( $P < 0.05$ ) with an obvious fold change ( $|\log_2FC| > 1$ ), of which 124 were upregulated and 107 were downregulated (*Figure 4*). A PPI network (*Figure S3*) was also constructed to reveal the association between proteins encoded by DEGs.

We also applied the same differential analysis method for the data from the GEO databases (GSE31210, GSE50081,



**Figure 2** Survival analysis: survival time analysis of 509 patients with smoking status.

and GSE68465) to identify DEGs between smokers and non-smokers (Figure S4). A total of 81 RNAs were differentially expressed in both the TCGA-LUAD and GEO datasets.

The expression of genes in LUAD cells has been linked to smoking-related transversion mutations in lung tumors. To identify tobacco smoke-inducible enhancers at the cellular level, we used the same method to analyze the expression profile of A549 human LUAD cells treated with DMSO or CSC for 2 weeks (GSE69770; Figure S5). A total of 21 genes (upregulated: *MALAT1*, *RSPO3*, *CPLX2*, *BAAT*, *CLDN2*, *TM4SF4*, *NTS*, and *VTN*; downregulated: *IGF2*, *HIST1H3F*, *HIST1H4C*, *HIST1H4D*, *HIST1H4E*, *HIST1H3B*, *HIST1H2AI*, *HIST1H1E*, *HIST1H2AM*, *HIST1H2BE*, *HIST1H3C*, *HIST1H1D*, and *HIST1H1B*) were both differentially expressed between smoking and non-smoking LUAD patients and between CSC- and DMSO-treated A549 cells, which more reliably revealed that changes in these genes were caused by smoking and their upregulation or downregulation might play a more important role.

### Gene functional analysis

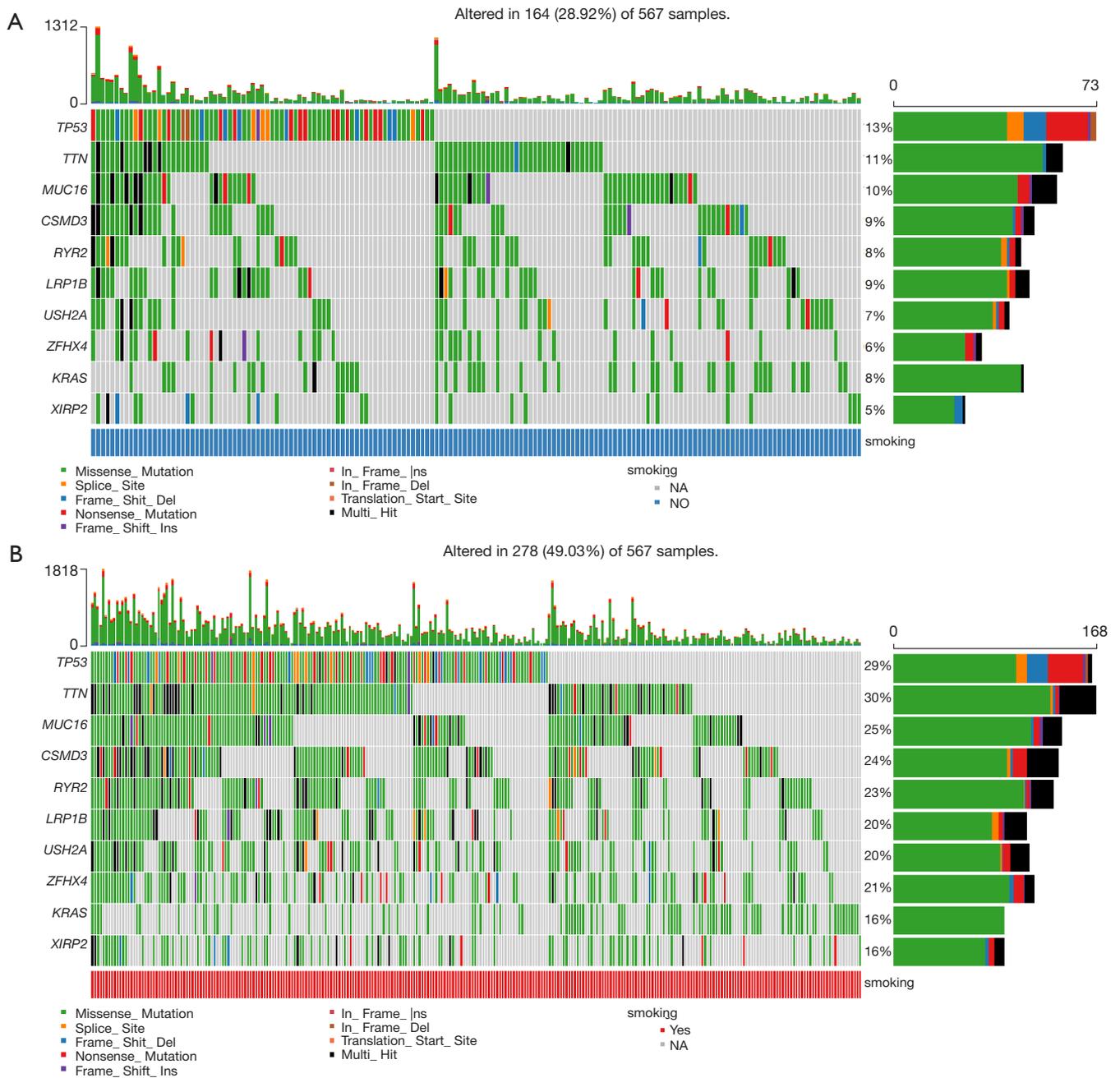
Gene set variation analysis (GSVA) of RNA sequencing data converts biological function annotations (GO and KEGG pathway analyses) into new expression matrices and treats them as general matrix files for differential

expression analysis. This analysis revealed 42 significant pathways between smokers and non-smokers (Figure 5); e.g., a Reactome for ethanol oxidation and the Amundson DNA damage response TP53 were downregulated and *Ly* aging middle DN, and Montero thyroid cancer poor survival up were upregulated in nonsmokers.

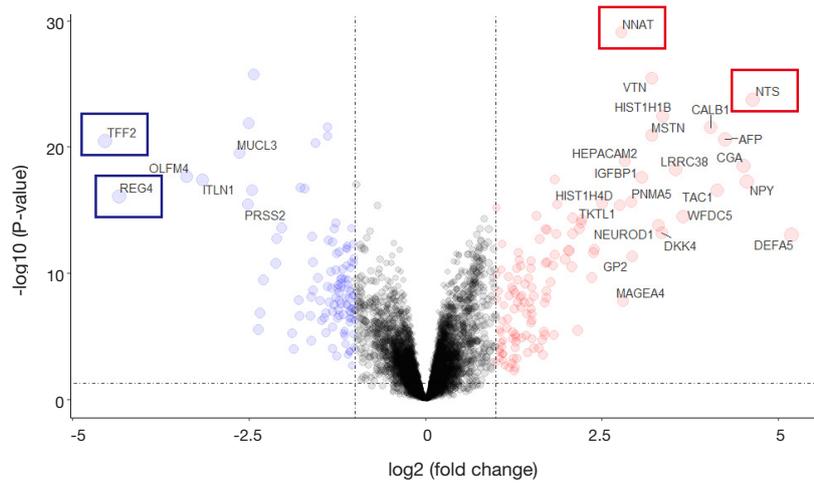
We also performed Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) analyses of the 231 genes that had differential somatic mutations or were differentially expressed based on the TCGA data. Gene functional enrichment analysis showed that six GO functional groups exhibited significant differences between smokers and non-smokers: extracellular matrix structural constituent, receptor ligand activity, structural constituent of cytoskeleton, actin filament binding, hormone activity, and receptor regulator activity (Figure S6A). For KEGG pathway analysis, four pathways differed between smokers and non-smokers: systemic lupus erythematosus, alcoholism, thyroid hormone synthesis, and protein digestion and absorption (Figure S6B).

### Differentially expressed miRNAs and functional annotation in TCGA-LUAD

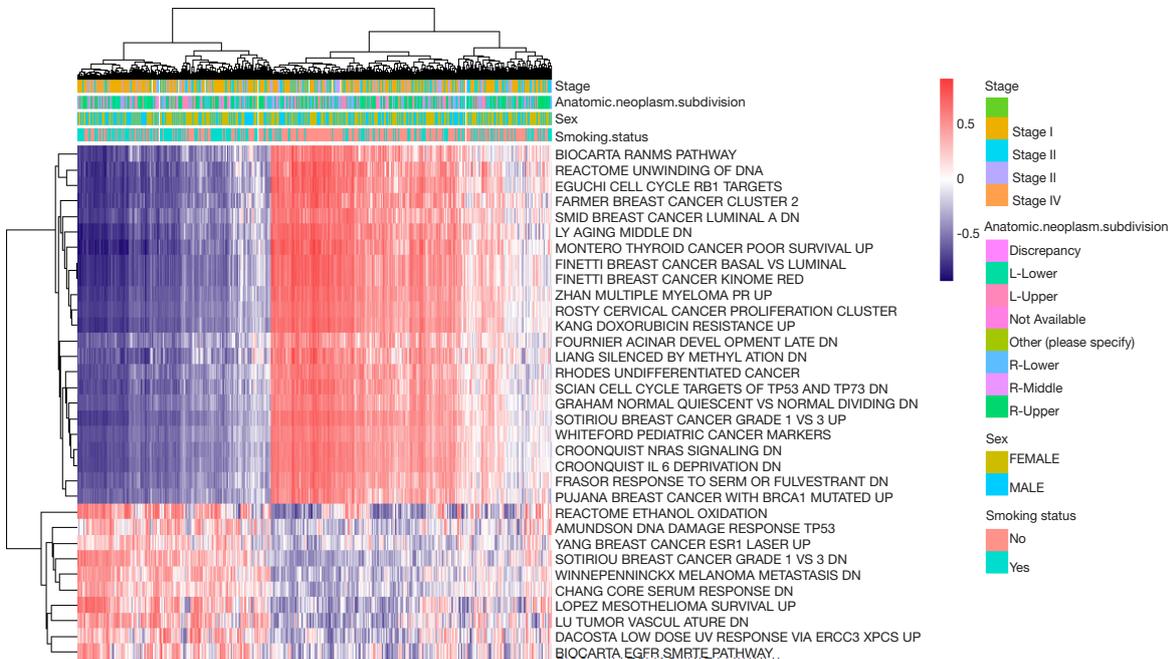
Of the 1,881 miRNAs expressed in the TCGA-LUAD with clinical information, 38 miRNAs were significantly upregulated ( $P < 0.05$ ,  $|\log_{2}FC| > 1$ ) and 7 were downregulated between smokers and non-smokers using



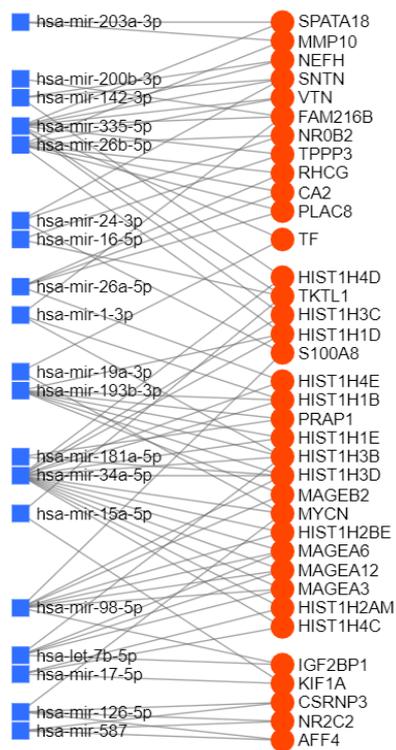
**Figure 3** Somatic mutation waterfall map grouped by smoking status, the red band below corresponded to the smoking group, and the blue was the non-smoking group.



**Figure 4** Differentially expressed genes in non-Smoking lung adenocarcinoma in TCGA-LUAD dataset. Volcanic map reviewed genes differentially expressed between the smoker and non-smoker groups. Blue dots represented significantly down-regulated genes, red dots represented significantly up-regulated genes, and black dots represented genes that are not differentially expressed. TCGA, The Cancer Genome Atlas; LUAD, lung adenocarcinoma.



**Figure 5** Heatmap of gene set variation analysis for and RNA-Seq data (GSVA).



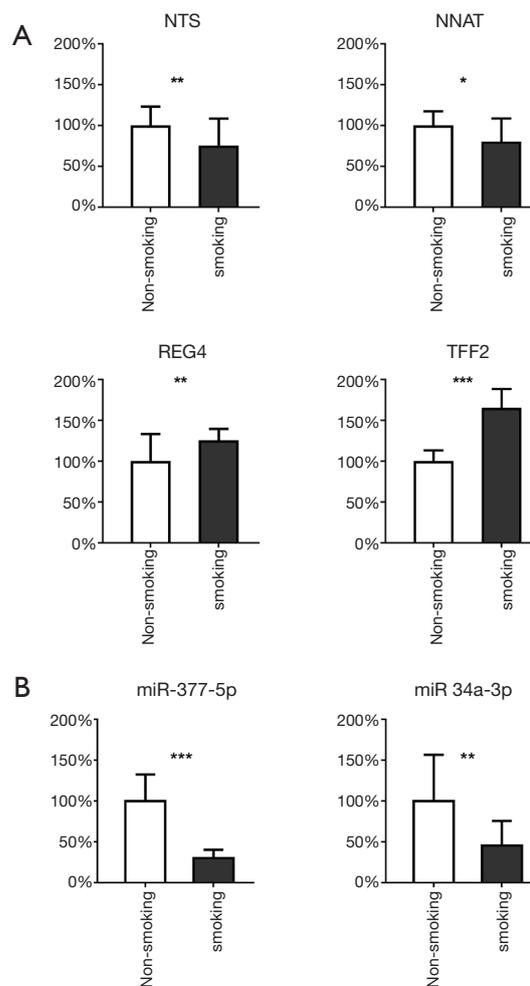
**Figure 6** Seventy-nine microRNAs-differentially expressed gene (miRNA-DEG) pairs consisting of 19 miRNAs and 36 mRNAs by miRNet.

the edgeR algorithm (Figure S7).

To determine the regulatory relationship between DEGs and miRNAs, we looked for links between the genes and miRNAs using the miRNet platform and obtained 7,945 pairs of miRNA-DEGs, while in consideration of the appearance of the figure, only the closest 85 pairs were shown in Figure 6 (differential genes were defined as  $P < 0.001$  and  $|\log FC| > 1$  and differential miRNAs were defined as  $P < 0.05$ ). miR-335-5p and miR-34a-5p are located at the core of the network, suggesting these miRNAs and their binding genes may play an important role in the differences between smoking and non-smoking lung cancer mechanisms.

#### Validation of mRNA and miRNA expression by qRT-PCR

After comprehensively considering the significance



**Figure 7** Validation of comparatively more significantly different mRNAs and miRNAs expression by qRT-PCR. \*,  $P < 0.05$ ; \*\*,  $P < 0.01$ ; \*\*\*,  $P < 0.005$ .

of differences and the FC value, we selected 4 highly significantly differential genes and 2 miRNAs. NTS, NNAT and miR-377-5p, miR-34a-3p are the most up-regulated factors in non-smoking adenocarcinoma and the TFF2 and REG4 are the most down-regulated in TCGA data as shown in Figure 4 and Figure S7. Therefore, we used qRT-PCR to measure the expression levels of them in 20 pairs of non-smoking LUAD and smoking adenocarcinoma samples. As shown in Figure 7, the alteration of the factors' expression in the results of qRT-PCR was generally consistent with that shown in the bioinformatics

analyses though the FCs were not as large as those in the bioinformatics results.

### Microenvironmental differences

Tumor immune cell infiltration refers to the migration of immune cells from the blood to the tumor tissue, where they exert their function. Infiltrating immune cells can be isolated from the tumor tissue. Resting mast cells, M2 macrophages, memory resting CD4<sup>+</sup> T cells, and resting dendritic cells were upregulated in non-smokers, while CD8<sup>+</sup> T cells, plasma cells, activated CD4<sup>+</sup> T cells, M1 macrophages, and follicular helper T cells were downregulated (*Figure 8A*). Furthermore, we looked for differences in immune-related gene expression. Nine genes related to antimicrobials (*Figure 8B*), six genes related to cytokines (*Figure 8C*), granzyme B (GZMB) for natural killer (NK) cell cytotoxicity (*Figure 8D*), and nuclear receptor subfamily 0 group B member 2 (NR0B2) for cytokine receptors (*Figure 8E*) were found to be significantly differentially expressed, indicating that cytotoxic and antimicrobial function may be one reason for the differences in the immune microenvironment between smokers and non-smokers. For programmed death-ligand 1 (PD-L1) (CD274), a significant difference in expression was also observed between smokers and non-smokers (*Figure 8F*), but with a low fold change value (logFC = 0.09).

The stemness indices (mRNAsi) were calculated for the TCGA data. mRNAsi is an index calculated based on expression data, which ranges from 0 to 1; a value close to 1 indicates a lower degree of cell differentiation and stronger stem cell characteristics (16). We found a significant difference in stemness indices between the groups (496 samples; 207 non-smokers and 289 smokers,  $P < 0.05$ , *Figure 9*), which confirmed the original results that non-smokers or long-term reformed smokers exhibit lower stemness than current and recently reformed smokers, as previously reported (16).

### Discussion

The study of cancer predisposition and cancer-specific genes is an important research direction. In this study, we observed quantitative differences between the LUAD genome of never smokers compared with smokers. Many studies before concluded and verified that smoke would change human bronchial and pulmonary alveolar epithelial cells via affecting mitochondrial function and regulating the

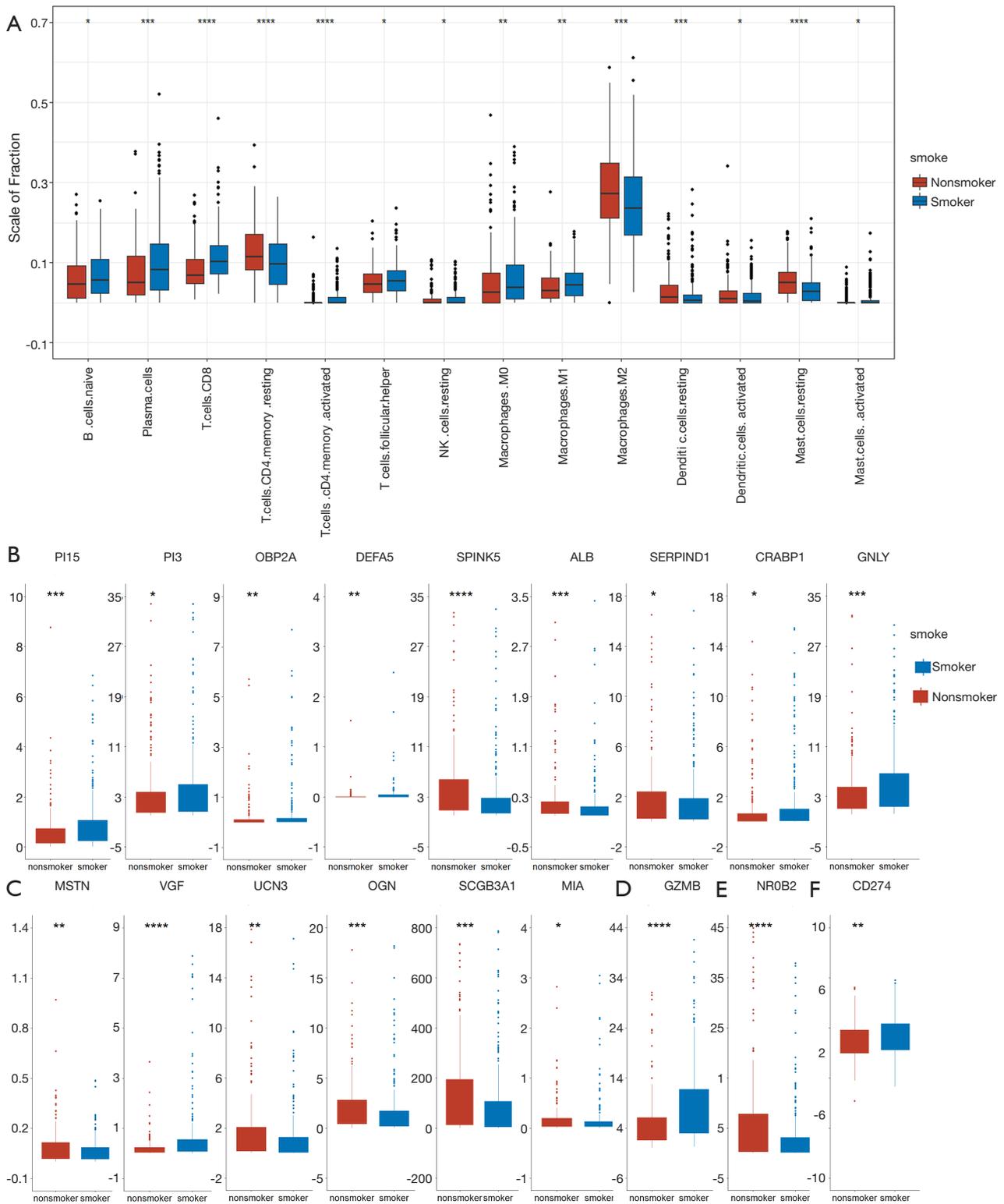
expression of several genes, especially PP2A, which might be connected with oncogenesis (17-20). However, these studies still focused more on how smoking affects lung acinar epithelial cells and causes chronic lung disease rather than tumors.

We proved that these differences existed at different levels of the genome, including somatic mutations, RNAs, and miRNAs, and some, mainly RNAs, showed significant increases or decreases at the cellular level.

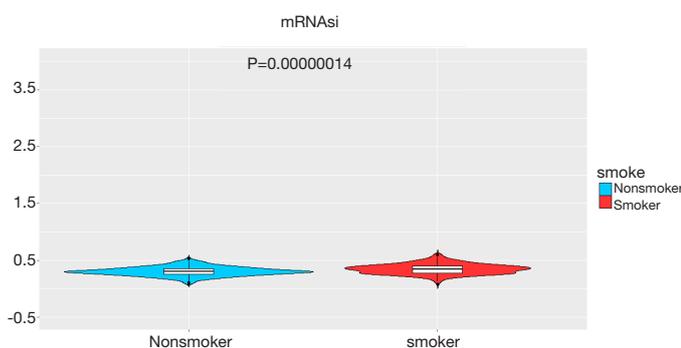
Boeckx *et al.* (21) found that the frequency of EGFR mutations is higher, and the frequency of TP53 mutations, is reduced in never smokers compared with smokers. However, in our study, these genes only showed differences at the somatic level, and not the RNA level, which is in contrast to previous studies of smoking-related genomic and/or transcriptional alterations. These differences may be due to the selection of different patient populations, tumor characteristics, and cohort sizes (22). In our previous study (23), we found that over half of our patients carried EGFR mutations, followed by KRAS and HER2 mutations, the first two of which also show different rate of mutation in this study, while HER2 rem ALK, RET, and ROS1 translocations, PIK3CA, BRAF, and NRAS mutations were rare and all were identified in less than 2% of patients. PD-1 positive tumors were identified in 14 (1.9%) patients, and PD-L1 positive tumors in 95 (13.0%) patients.

In the present study, NTS and NNAT were upregulated in non-smoking LUAD patients, DMSO-treated A549 LUAD cells and our own samples. NTS, along with its receptor NTSR, has been shown to be important in lung cancer outcomes (24,25) and therapies against NTS can decrease tumor growth and metastasis (26,27). These data suggested that the overexpression of NTS in non-smokers may be a trigger for the development of LUAD. NNAT mRNA is mainly expressed in endocrine and adipose tissues in a hormone- and nutrient-sensitive manner (28-30) and has the potential to be used as a marker differentiating between Large-Cell Neuroendocrine Carcinoma (LCNEC) and small cell lung cancer (SCLC) as it more highly expressed in LCNEC, making it different from smoker NSCLC LUAD (31,32).

TFF2 and REG4 were both downregulated in non-smoking LUAD in TCGA and our own samples. Lung macrophages rely upon TFF2 to promote epithelial proliferation, while in the absence of TFF2, lung epithelia were unable to proliferate and expressed reduced lung mRNA transcript levels (33,34). In non-smoking LUAD patients, down-regulation of TFF2 might indicate weaker



**Figure 8** Differences of microenvironment in non-smoking lung adenocarcinoma. (A) Comparison of each leukocyte fraction between smokers and nonsmokers; (B) differential genes related to antimicrobials; (C) Differential genes related to cytokines; (D) differential genes related to natural killer (NK) cell cytotoxicity; (E) differential gene related to cytokine receptors; (F) differential gene related to PD-L1. PD-L1, programmed death-ligand 1. \*, P<0.05; \*\*, P<0.01; \*\*\*, P<0.005; \*\*\*\*, P<0.001.



**Figure 9** Differences of stemness between smokers and non-smokers in LUAD. LUAD, lung adenocarcinoma.

lung repair ability, which promotes tumor development. The overexpression of REG4, which was shown in the smoking group previously, was proven to be closely correlated to the carcinogenesis in some types of cancer (35). Also, REG4 plays an important role in KRAS-driven lung cancer pathogenesis and is a novel biomarker of LUAD subtype (36). However, whether its expression can be used to distinguish non-smoking from smoking LUAD needs further experiments to verify, as it showed unstable CT values when verifying.

A group of histones including HIST1H3F, HIST1H4C, HIST1H4D, HIST1H4E, HIST1H3B, HIST1H2AI, HIST1H1E, HIST1H2AM, HIST1H2BE, HIST1H3C, HIST1H1D, and HIST1H1B differentially expressed between smokers and nonsmokers in our study. Histone variants act as transcriptional activators or repressors of cancer-related genes. Several researches about HIST1H2, HIST1H3, HIST1H4 found that they were related to the progression and metastasis of tumor cell (37-39). However, there reports on how smoking regulates the expression of these histones so far.

Overexpression of miR-377-5p, which was more often in the non-smokers in our study, was reported remarkably downregulated in NSCLC tissues, and inhibited cell proliferation and development as a tumor-suppressor gene (40,41). miR-34a overexpressed among the nonsmokers. Previous study found that p53 regulates PDL1 via miR-34 and SART3 overexpression increased miR-34a levels, which may affect cell cycle progression in NSCLC cells, while miR-34a inhibits NSCLC tumor growth and metastasis through targeting EGFR (42-44). Furthermore, it was

also considered to be a target in the therapy of lung cancer (45,46).

We also focused on gene function. The leukocyte fraction varied substantially across the immune subtypes (15). We found that CD8<sup>+</sup> T cells, plasma cells, activated CD4<sup>+</sup> T cells, M1 macrophages, and follicular helper T cells were comparatively downregulated in non-smokers. Previously Kinoshita *et al.* found a high FOXP3/CD4 ratio in smokers with adenocarcinoma, a low number of CD20<sup>+</sup> B cells in non-smokers was identified as an independent unfavorable prognostic factor in resected NSCLC, and infiltrating CD8<sup>+</sup> T cells may not be activated sufficiently in the immunosuppressive microenvironment in non-smokers with adenocarcinoma (47,48). Li *et al.* (49) concluded that aberrant activation of mast cells and CD4<sup>+</sup> memory T cells plays a crucial role in cigarette smoking-induced immune dysfunction in the lung, leading to tumor development and progression. Generally, tobacco smoking patients with NSCLC showed a higher PD-L1 tumor proportion score (50-52) and experience a higher response rate with immunotherapy, mainly pembrolizumab and nivolumab, than non-smokers (50,53-57). In the present study, we also found a significant difference in expression of the PD-L1 gene (CD274) between the two groups but only with a small fold-change value. All of this evidence suggested that smoking created a different immune environment with abnormal immune cells. Smokers appeared to develop LUAD more easily because their immune cells are not as limited as those of non-smokers.

A strong association between mRNAi and tobacco smoking status in LUAD was shown in this study. This confirmed previous research, which suggested that the stemness of LUAD tumors might be activated in response to environmental stimuli such as smoking, and might influence tumor aggressiveness (16).

There are some limitations to this study. Smoking status data in the TCGA database did not include information regarding whether the patients were frequently exposed to second-hand smoke or kitchen fumes, which are also important causes of lung cancer. Also, our research results were mainly based on the data of TCGA and verified with GEO datasets. We were unable to verify our results with other databases because of a lack of high-quality data on smoking. Lung cancer caused by smoking is a long-term process, so it is difficult to determine the role of these factors. Additional studies will be necessary to further

explore and verify these differences.

## Conclusions

In summary, our study characterized thorough patterns of genome alterations and the microenvironment in non-smoking LUAD. Compared with smokers, non-smoking LUAD patients had a lower mutation rate in somatic cells, RNAs, and miRNAs with a different microenvironment, including immune cell infiltration and stemness, which revealed the complex association between molecular mechanisms and clinical outcomes. We believe this study will improve our understanding of the mechanism underlying non-smoking LUAD, and effectively prevent, diagnose, and treat LUAD.

## Acknowledgments

The authors thank International Science Editing (<http://www.internationalscienceediting.com>) for editing this manuscript.

*Funding:* This work was supported by FDUROP (Fudan's Undergraduate Research Opportunities Program), FDUROP (19028), National University Student Innovation Program Zhengyi Scholar Foundation of School of Basic Medical Sciences, Fudan University (S22-11) and the Training Program for the Talents of Zhongshan Hospital, Fudan University (2019ZSYXGG06).

## Footnote

*Reporting Checklist:* The authors have completed the MDAR reporting checklist. Available at <http://dx.doi.org/10.21037/tlcr-20-276>

*Conflicts of Interest:* All authors have completed the ICMJE uniform disclosure form (available at <http://dx.doi.org/10.21037/tlcr-20-276>). The authors have no conflicts of interest to declare.

*Ethical Statement:* The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. All procedures performed in this study were in accordance with the Declaration of Helsinki (as revised in 2013) and was approved by the Ethics Committee of Zhongshan Hospital, Fudan University, China (B2019 – 377R). Because of the

retrospective nature of the research, the requirement for informed consent was waived.

*Open Access Statement:* This is an Open Access article distributed in accordance with the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License (CC BY-NC-ND 4.0), which permits the non-commercial replication and distribution of the article with the strict proviso that no changes or edits are made and the original work is properly cited (including links to both the formal publication through the relevant DOI and the license). See: <https://creativecommons.org/licenses/by-nc-nd/4.0/>.

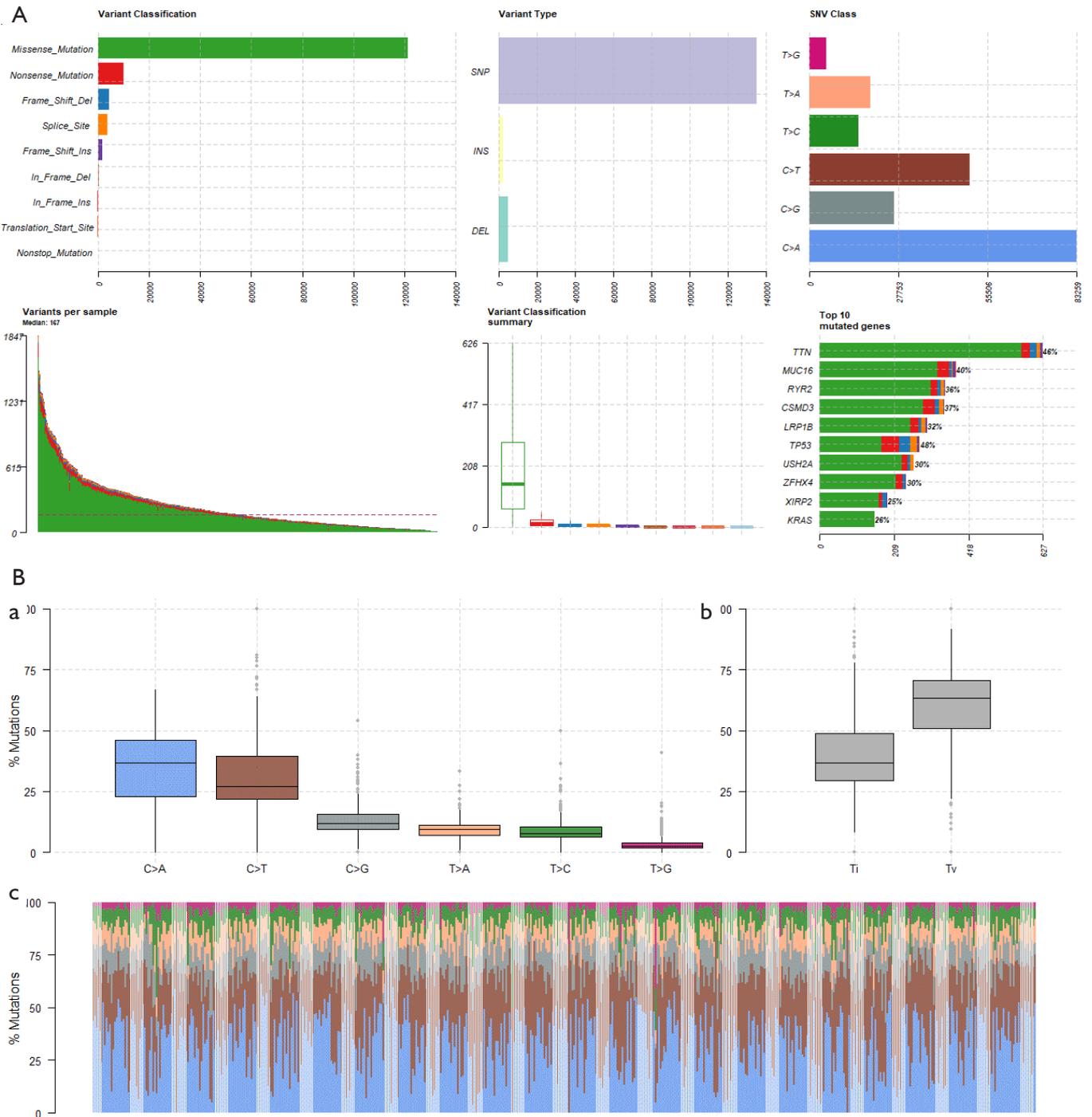
## References

1. Bray F, Ferlay J, Soerjomataram I, et al. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* 2018;68:394-424.
2. Yoon JY, Sigel K, Martin J, et al. Evaluation of the Prognostic Significance of TNM Staging Guidelines in Lung Carcinoid Tumors. *J Thorac Oncol* 2019;14:184-92.
3. WHO report on the global tobacco epidemic 2019. Available online: <https://apps.who.int/iris/bitstream/handle/10665/326043/9789241516204-eng.pdf?ua=1>, accessed 12 December 2019; Geneva: World Health Organization 2019.
4. WHO highlights huge scale of tobacco-related lung disease deaths. In: World No Tobacco Day 2019: Don't let tobacco take your breath away. Geneva: World Health Organization. Available online: <https://www.who.int/news-room/detail/29-05-2019-who-highlights-huge-scale-of-tobacco-related-lung-disease-deaths>. Accessed 12/12 2019.
5. Glantz S, Gonzalez M. Effective tobacco control is key to rapid progress in reduction of non-communicable diseases. *Lancet* 2012;379:1269-71.
6. Doll R, Hill AB. The mortality of doctors in relation to their smoking habits; a preliminary report. *Br Med J* 1954;1:1451-5.
7. Parkin DM, Bray F, Ferlay J, et al. Global Cancer Statistics, 2002. *CA Cancer J Clin* 2005;55:74-108.
8. Sun S, Schiller JH, Gazdar AF. Lung cancer in never smokers — a different disease. *Nat Rev Cancer* 2007;7:778-90.
9. Toh CK, Gao F, Lim WT, et al. Never-Smokers With Lung Cancer: Epidemiologic Evidence of a Distinct Disease Entity. *J Clin Oncol* 2006;24:2245-51.
10. Akhtar N, Bansal JG. Risk factors of Lung Cancer in

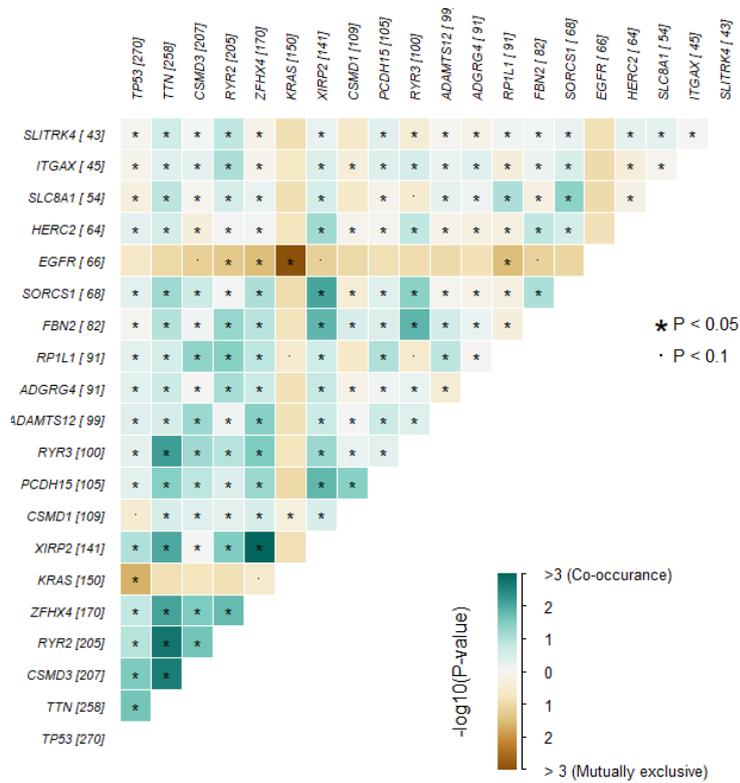
- nonsmoker. *Curr Probl Cancer* 2017;41:328-39.
11. Der SD, Sykes J, Pintilie M, et al. Validation of a histology-independent prognostic gene signature for early-stage, non-small-cell lung cancer including stage IA patients. *J Thorac Oncol* 2014;9:59-64.
  12. Shedden K, Taylor JMG, Enkemann SA, et al. Gene expression-based survival prediction in lung adenocarcinoma: a multi-site, blinded validation study. *Nat Med* 2008;14:822-7.
  13. Okayama H, Kohno T, Ishii Y, et al. Identification of genes upregulated in ALK-positive and EGFR/KRAS/ALK-negative lung adenocarcinomas. *Cancer Res* 2012;72:100-11.
  14. Stueve TR, Li WQ, Shi J, et al. Epigenome-wide analysis of DNA methylation in lung tissue shows concordance with blood studies and identifies tobacco smoke-inducible enhancers. *Hum Mol Genet* 2017;26:3014-27.
  15. Thorsson V, Gibbs DL, Brown SD, et al. The Immune Landscape of Cancer. *Immunity* 2018;48:812-30.e14.
  16. Malta TM, Sokolov A, Gentles AJ, et al. Machine Learning Identifies Stemness Features Associated with Oncogenic Dedifferentiation. *Cell* 2018;173:338-54.e15.
  17. Doherty DF, Nath S, Poon J, et al. Protein Phosphatase 2A Reduces Cigarette Smoke-induced Cathepsin S and Loss of Lung Function. *Am J Respir Crit Care Med* 2019;200:51-62.
  18. Liu CY, Huang TT, Chen YT, et al. Targeting SET to restore PP2A activity disrupts an oncogenic CIP2A-feedforward loop and impairs triple negative breast cancer progression. *EBioMedicine* 2019;40:263-75.
  19. Ballweg K, Mutze K, Königshoff M, et al. Cigarette smoke extract affects mitochondrial function in alveolar epithelial cells. *Am J Physiol Lung Cell Mol Physiol* 2014;307:L895-L907.
  20. Mizumura K, Cloonan SM, Nakahira K, et al. Mitophagy-dependent necroptosis contributes to the pathogenesis of COPD. *J Clin Invest* 2014;124:3987-4003.
  21. Boeckx B, Shahi RB, Smeets D, et al. The genomic landscape of nonsmall cell lung carcinoma in never smokers. *Int J Cancer* 2020;146:3207-18.
  22. Karlsson A, Ringnér M, Lauss M, et al. Genomic and Transcriptional Alterations in Lung Adenocarcinoma in Relation to Smoking History. *Clin Cancer Res* 2014;20:4912.
  23. Zhao M, Zhan C, Li M, et al. Aberrant status and clinicopathologic characteristic associations of 11 target genes in 1,321 Chinese patients with lung adenocarcinoma. *J Thorac Dis* 2018;10:398-407.
  24. Reinecke M. Neurotensin. Immunohistochemical localization in central and peripheral nervous system and in endocrine cells and its functional role as neurotransmitter and endocrine hormone. *Prog Histochem Cytochem* 1985;16:1-172.
  25. Alifano M, Souazé F, Dupouy S, et al. Neurotensin receptor 1 determines the outcome of non-small cell lung cancer. *Clin Cancer Res* 2010;16:4401-10.
  26. Younes M, Wu Z, Dupouy S, et al. Neurotensin (NTS) and its receptor (NTSR1) causes EGFR, HER2 and HER3 over-expression and their autocrine/paracrine activation in lung tumors, confirming responsiveness to erlotinib. *Oncotarget* 2014;5:8252-69.
  27. Wu Z, Fournel L, Stadler N, et al. Modulation of lung cancer cell plasticity and heterogeneity with the restoration of cisplatin sensitivity by neurotensin antibody. *Cancer Lett* 2019;444:147-61.
  28. Millership SJ, Da Silva Xavier G, Choudhury AI, et al. Neuronatin regulates pancreatic cell insulin content and secretion. *J Clin Invest* 2018;128:3369-81.
  29. Usui H, Morii K, Tanaka R, et al. cDNA cloning and mRNA expression analysis of the human neuronatin. High level expression in human pituitary gland and pituitary adenomas. *J Mol Neurosci* 1997;9:55-60.
  30. Joseph R, Dou D, Tsang W. Molecular cloning of a novel mRNA (neuronatin) that is highly expressed in neonatal mammalian brain. *Biochem Biophys Res Commun* 1994;201:1227-34.
  31. Nass N, Walter S, Jechorek D, et al. High neuronatin (NNAT) expression is associated with poor outcome in breast cancer. *Virchows Arch* 2017;471:23-30.
  32. Okubo C, Minami Y, Tanaka R, et al. Analysis of differentially expressed genes in neuroendocrine carcinomas of the lung. *J Thorac Oncol* 2006;1:780-6.
  33. Hung LY, Oniskey TK, Sen D, et al. Trefoil Factor 2 Promotes Type 2 Immunity and Lung Repair through Intrinsic Roles in Hematopoietic and Nonhematopoietic Cells. *Am J Pathol* 2018;188:1161-70.
  34. Hung LY, Sen D, Oniskey TK, et al. Macrophages promote epithelial proliferation following infectious and non-infectious lung injury through a Trefoil factor 2-dependent mechanism. *Mucosal Immunol* 2019;12:64-76.
  35. Jin J, Lv H, Wu J, et al. Regenerating Family Member 4 (Reg4) Enhances 5-Fluorouracil Resistance of Gastric Cancer Through Activating MAPK/Erk/Bim Signaling Pathway. *Med Sci Monit* 2017;23:3715-21.
  36. Sun S, Hu Z, Huang S, et al. REG4 is an indicator for

- KRAS mutant lung adenocarcinoma with TTF-1 low expression. *J Cancer Res Clin Oncol* 2019;145:2273-83.
37. Park SM, Choi EY, Bae M, et al. Histone variant H3F3A promotes lung cancer cell migration through intronic regulation. *Nat Commun* 2016;7:12914.
  38. Ju J, Chen A, Deng Y, et al. NatD promotes lung cancer progression by preventing histone H4 serine phosphorylation to activate Slug expression. *Nat Commun* 2017;8:928.
  39. Hsu CC, Zhao D, Shi J, et al. Gas41 links histone acetylation to H2A.Z deposition and maintenance of embryonic stem cell identity. *Cell Discov* 2018;4:28.
  40. Wu H, Liu HY, Liu WJ, et al. miR-377-5p inhibits lung cancer cell proliferation, invasion, and cell cycle progression by targeting AKT1 signaling. *J Cell Biochem* 2018. doi: 10.1002/jcb.28091.
  41. Zhang J, Zhao M, Xue ZQ, et al. miR-377 inhibited tumorous behaviors of non-small cell lung cancer through directly targeting CDK6. *Eur Rev Med Pharmacol Sci* 2016;20:4494-9.
  42. Cortez MA, Ivan C, Valdecanas D, et al. PDL1 Regulation by p53 via miR-34. *J Natl Cancer Inst* 2015;108:djv303.
  43. Sherman EJ, Mitchell DC, Garner AL. The RNA-binding protein SART3 promotes miR-34a biogenesis and G cell cycle arrest in lung cancer cells. *J Biol Chem* 2019;294:17188-96.
  44. Li YL, Liu XM, Zhang CY, et al. MicroRNA-34a/EGFR axis plays pivotal roles in lung tumorigenesis. *Oncogenesis* 2017;6:e372.
  45. Xue W, Dahlman JE, Tammela T, et al. Small RNA combination therapy for lung cancer. *Proc Natl Acad Sci U S A* 2014;111:E3553-61.
  46. Jiang ZQ, Li MH, Qin YM, et al. Luteolin Inhibits Tumorigenesis and Induces Apoptosis of Non-Small Cell Lung Cancer Cells via Regulation of MicroRNA-34a-5p. *Int J Mol Sci* 2018;19:447.
  47. Kinoshita T, Kudo-Saito C, Muramatsu R, et al. Determination of poor prognostic immune features of tumour microenvironment in non-smoking patients with lung adenocarcinoma. *Eur J Cancer* 2017;86:15-27.
  48. Kinoshita T, Muramatsu R, Fujita T, et al. Prognostic value of tumor-infiltrating lymphocytes differs depending on histological type and smoking habit in completely resected non-small-cell lung cancer. *Ann Oncol* 2016;27:2117-23.
  49. Li X, Li J, Wu P, et al. Smoker and non-smoker lung adenocarcinoma is characterized by distinct tumor immune microenvironments. *Oncoimmunology* 2018;7:e1494677.
  50. Gainor JF, Shaw AT, Sequist LV, et al. EGFR Mutations and ALK Rearrangements Are Associated with Low Response Rates to PD-1 Pathway Blockade in Non-Small Cell Lung Cancer: A Retrospective Analysis. *Clin Cancer Res* 2016;22:4585-93.
  51. Rangachari D, VanderLaan PA, Shea M, et al. Correlation between Classic Driver Oncogene Mutations in EGFR, ALK, or ROS1 and 22C3-PD-L1  $\geq 50\%$  Expression in Lung Adenocarcinoma. *J Thorac Oncol* 2017;12:878-83.
  52. Pan Y, Zheng D, Li Y, et al. Unique distribution of programmed death ligand 1 (PD-L1) expression in East Asian non-small cell lung cancer. *J Thorac Dis* 2017;9:2579-86.
  53. Mok TSK, Wu YL, Kudaba I, et al. Pembrolizumab versus chemotherapy for previously untreated, PD-L1-expressing, locally advanced or metastatic non-small-cell lung cancer (KEYNOTE-042): a randomised, open-label, controlled, phase 3 trial. *Lancet* 2019;393:1819-30.
  54. Sundar R, Soong R, Cho BC, et al. Immunotherapy in the treatment of non-small cell lung cancer. *Lung Cancer* 2014;85:101-9.
  55. Soria JC, Gettinger S, Gordon M, et al. 1322biomarkers associated with clinical activity of pd-1l blockade in non-small cell lung cancer (NSCLC) patients (pts) in a phase I study of MPDL3280A. *Ann Oncol* 2014;25:iv465.
  56. Gandhi L, Rodríguez-Abreu D, Gadgeel S, et al. Pembrolizumab plus Chemotherapy in Metastatic Non-Small-Cell Lung Cancer. *N Engl J Med* 2018;378:2078-92.
  57. Borghaei H, Paz-Ares L, Horn L, et al. Nivolumab versus Docetaxel in Advanced Nonsquamous Non-Small-Cell Lung Cancer. *N Engl J Med* 2015;373:1627-39.

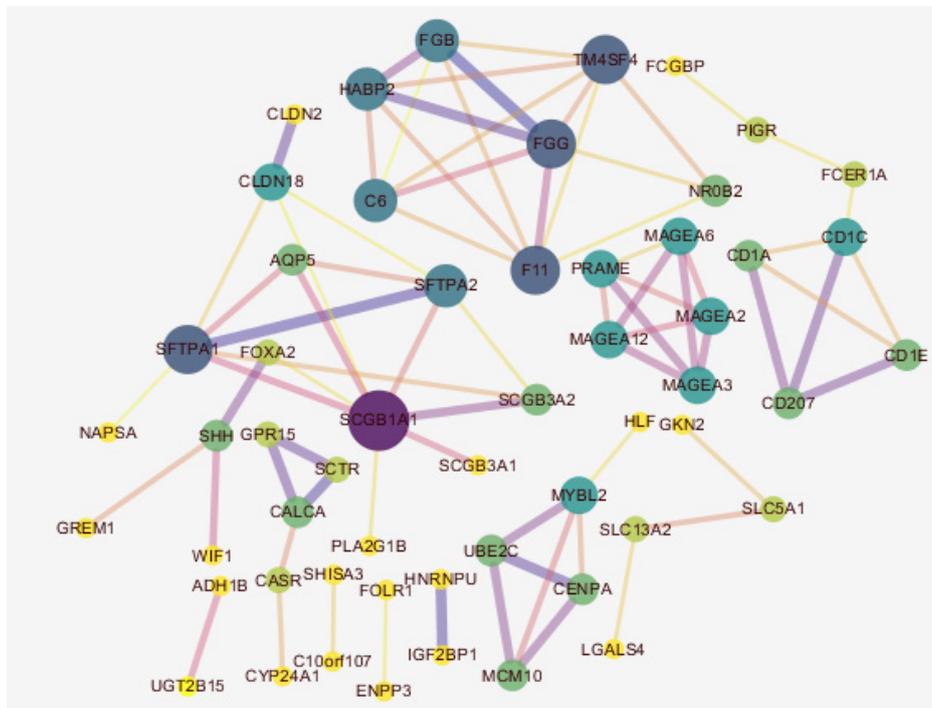
**Cite this article as:** Sui Q, Liang J, Hu Z, Chen Z, Bi G, Huang Y, Li M, Zhan C, Lin Z, Wang Q. Genetic and microenvironmental differences in non-smoking lung adenocarcinoma patients compared with smoking patients. *Transl Lung Cancer Res* 2020;9(4):1407-1421. doi: 10.21037/tlcr-20-276



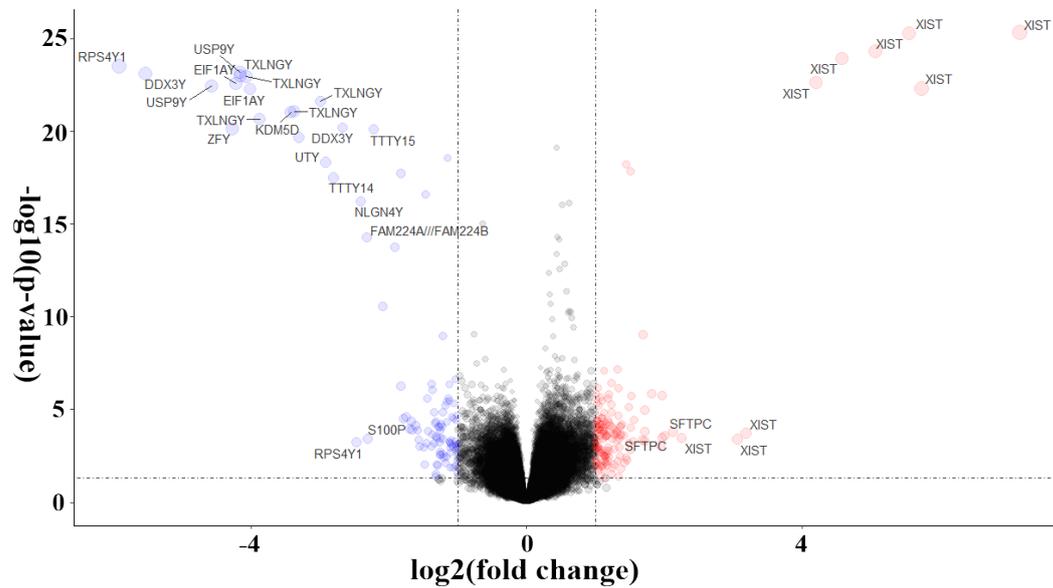
**Figure S1** The summary of the LUAD patients' somatic mutation data, (A) displayed number of variants in each sample as a stacked bar plot and variant types as a boxplot summarized; (B) classified SNPs into transitions and transversions, (a) showed the overall distribution of the six different transformations, (b) classified the SNPs as transitions (Ti) and transversions (Tv), showing their proportion, (c) stacked bar graph of the percent conversion in each sample.



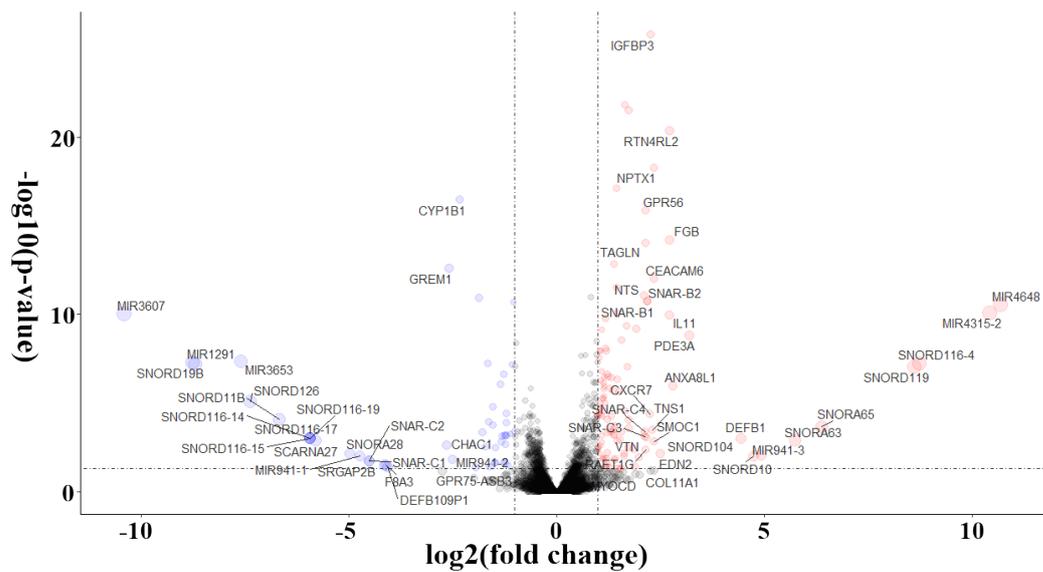
**Figure S2** Exclusive/co-occurrence event analysis on top 20 differently mutated genes.



**Figure S3** Protein-protein interaction (PPI) network of differently expressed genes between smokers and non-smokers in TCGA-LUAD. The size and gradient color of the Node were adjusted by degree, and the thickness and gradient color of edge are adjusted by combined score.

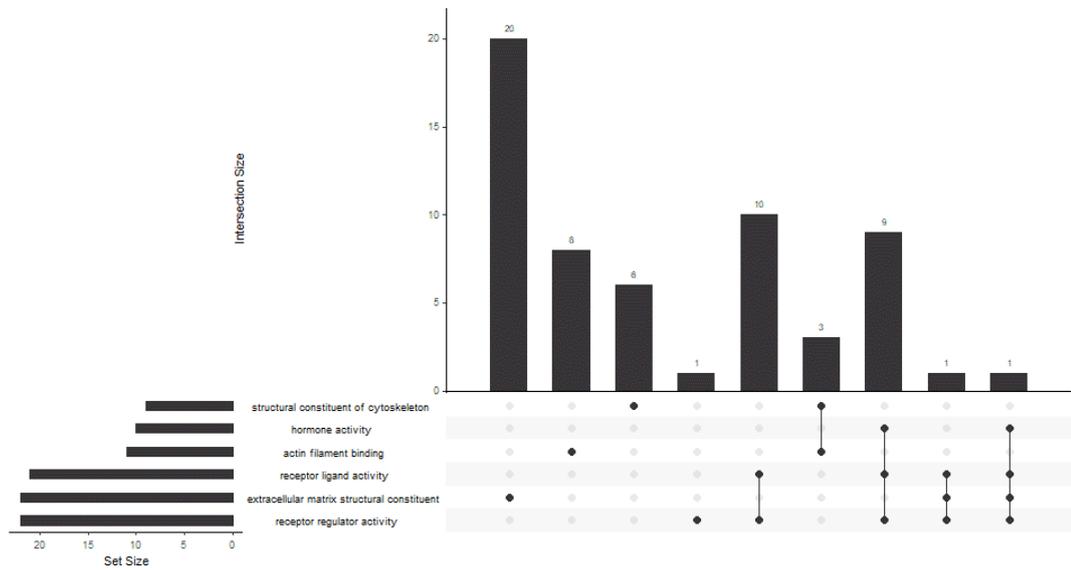


**Figure S4** Differentially expressed genes differentially expressed genes in non-smoking lung adenocarcinoma associated with smoking in GEO datasets. GEO, Gene Expression Omnibus.

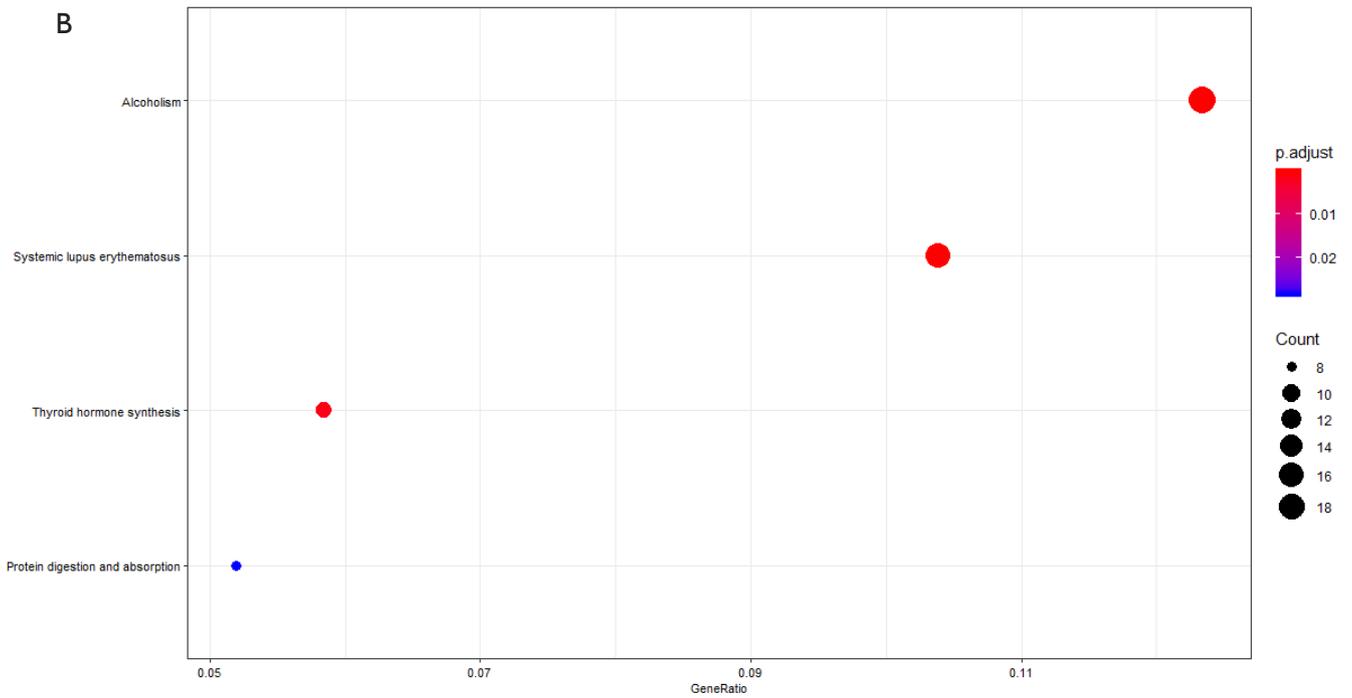


**Figure S5** Differentially expressed genes between A549 cells treated with DMSO and cigarette smoke condensate (CSC) in GSE69770 datasets.

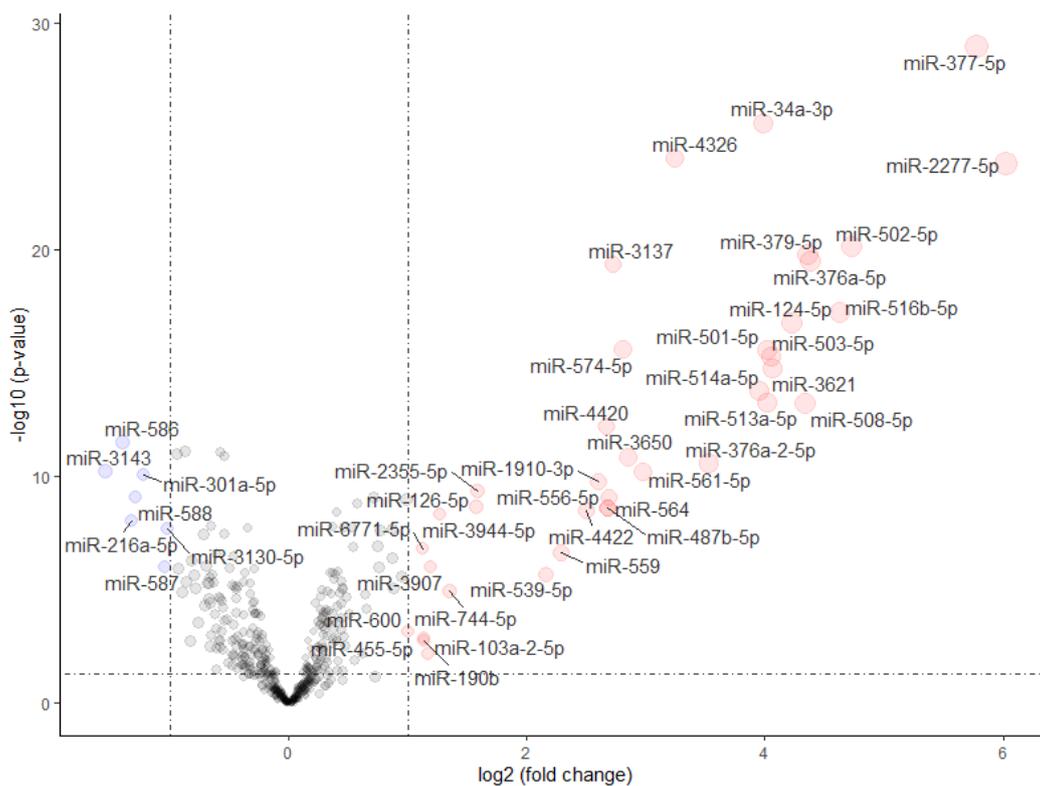
A



B



**Figure S6** Flip graph of the functional enrichment analysis of differential mutation and expression genes between smokers and non-smokers group which was more focused on overlapping of genes between different gene sets (A); bubble chart for all significantly different KEGG pathways (B).



**Figure S7** Differentially expressed miRNA in non-smoking lung adenocarcinoma in TCGA-LUAD dataset.

**Table S1** The sequences and melting temperature ( $T_m$ ) of the primers used in our research, whether they span exon junctions, PCR efficiency and correlation with dilution series ( $R^2$ )

Gene	Primers sequence 5'-3' (forward, reverse)	$T_m$ (°C)	Exon junction span	PCR efficiency (%)	Correlation with dilution series ( $R^2$ )
NTS	TGCTTTAGATGGCTTTAGCTTGG	59.9	Yes	101.8	0.998
	TTCCTGGATTA ACTCCCAGTGT	59.8			
NNAT	ACTGGGTAGGATTCGCTTTTCG	59.9	Yes	99.8	0.999
	ACACCTCACTTCTCGCAATGG	59.2			
TFF2	GCTGTTTCGACTCCAGTGTCA	64.0	Yes	101.5	0.999
	CCACAGTTTCTTCGGTCTGAG	60.2			
REG4	CTGCTCCTATTGCTGAGCTG	58.9	Yes	101.2	0.996
	GGACTTGTGGTAAACCATCCAG	58.6			