The benefits of smoking cessation on survival in cancer patients by integrative analysis of multi-omics data

CURRENT STATUS: Posted

Sheng Yang, Tong Liu, Geyu Liang

Sheng Yang
Southeast University

1024227634@qq.com Corresponding Author
ORCiD: https://orcid.org/0000-0002-0485-7728

Tong Liu
Southeast University

Geyu Liang
Southeast University

Prescreen

10.21203/rs.3.rs-28649/v1

Subject Areas

Cancer Biology

Keywords

current smokers, reformed smokers, smoking signature, prognosis
Abstract

**Background:** Few studies have examined the association between smoking cessation at the time of diagnosis and overall survival among cancer patients. We aimed to assess the benefits of quitting smoking on prognosis.

**Methods:** We obtained cohorts of smoking-related cancer patients from the Cancer Genome Atlas (TCGA) database. Hazard ratios (HR) with 95% confidence intervals (CIs) were calculated to evaluate smoking behavior at cancer diagnosis (reformed smokers vs. current smokers) in association with overall survival using age and multivariate-adjusted Cox regressions analysis. Comprehensive analysis of multi-component data was carried out to determine the genetic and epigenetic landscape differences and their effects on the differential expression genes (DEGs) between reformed smokers and current smokers. The smoking signature was constructed using univariate Cox regression and LASSO-Cox regression model.

**Results:** According to the multivariate-adjusted Cox regressions analysis, quitting smoking was the independent protective factor for overall survival in lung squamous cell carcinoma (LUSC) (HR=0.67, 95%CI=0.48-0.94). A total of 85 key DEGs were found to be affected by different modes of genetic and epigenetic regulation, which might represent key drivers in smokers. At last, we provided the smoking signature which could predict prognosis with high power (HR=1.70, 95% CI=1.19-2.43, AUC= 0.65, 0.67 and 0.70 for 2, 3 and 5-year survival, respectively). The smoking signature was also applicable in other smoking-related cancers, including bladder urothelial carcinoma (HR=1.70, 95% CI=1.01-2.88), cervical carcinoma (HR=5.69, 95% CI=1.37-23.69), head and neck squamous cell carcinoma (HR=1.97, 95% CI=1.41-2.76), lung adenocarcinoma (HR=1.73, 95% CI=1.16-2.57), and pancreatic adenocarcinoma (HR=4.28, 95% CI=1.47-12.47).

**Conclusions:** The present study demonstrated that quitting smoking at diagnosis decreased risk of death in cancer patients. We also provided the smoking signature to evaluate the true effect of smoking, suggesting that smoking cessation could be a part of cancer treatment to improve the survival rate of cancer patients.

Background

Tobacco smoking is a risk factor for the occurrence and increases the incidence of various cancers, including bladder(1), head and neck(2), lung(3), and pancreatic cancer(4). Tobacco smoking contains many carcinogenic chemicals that can create a specific mutational signature and increase the somatic mutational burden associated with unrepaired DNA damage(5). In addition to causing frequent gene mutations, tobacco smoking also appears to break the immune homeostasis, which may contribute to tumorigenesis(6)). It is recognized that continuous smoking not only has unhealthy impacts on the general population, but also is the negative prognostic factor for cancer patients by comparing smokers with never smokers in most studies(7,8). However, few studies have researched the association between smoking status changes (especially quitting smoking) and mortality among cancer patients compared with the general population. A study found that current smoking increased overall mortality risk compared with former smokers using multivariate Cox regression analysis(9). Another study based on the Shanghai Cohort Study also found that a statistically significant increased mortality risk was associated with smoking relative to nonsmoking after cancer diagnosis(10). A cohort in Japan also found that quitters had consistently higher survival rates than current smokers during a ten-year calendar period after diagnosis among cancer patients and suggested that smoking cessation should be a part of cancer care(11). These studies were based on a large population and adjusted for age, gender, stage and other basic characters. However, some important prognostic factors were not considered, including human papillomavirus (HPV) status associated with the prognosis of head and neck cancer and cervical cancer(12), and tumor status after surgery associated with the prognosis in many cancers(13,14). It is another limit that the underlying mechanism of smoking cessation to improve survival time has not been further studied. Therefore, to understand the benefits of quitting smoking on prognosis among cancer patients, we first evaluated smoking status at cancer diagnosis (reformed smokers vs. current smokers) in association with overall survival and then comprehensively analyzed the transcriptome data, mutational profile and immune microenvironment of smoking-related cancers from the Cancer Genome Atlas (TCGA) database.
Materials And Methods

Data source

This study made use of public data from the TCGA database. The information of smoking status, survival time and the clinical characteristics were downloaded for bladder urothelial carcinoma (BLCA), cervical squamous cell carcinoma and endocervical adenocarcinoma (CESC), esophageal carcinoma (ESCA), head and neck squamous cell carcinoma (HNSC), lung adenocarcinoma (LUAD), lung squamous cell carcinoma (LUSC) and pancreatic adenocarcinoma (PAAD) from TCGA data portal (https://portal.gdc.cancer.gov/). Detailed patient characteristics of each cancer were given in Supplementary table 1. Besides, the gene expression RNAsq (HTSeq-FPKM), miRNA expression RNAsq (IlluminaHiSeq), somatic mutation data (SNV, VarScan2 Variant Aggregation and Masking), copy number variation data (CNV, Masked Copy Number Segment hg38) and DNA methylation data (Illumina Human Methylation 450) of above cancers were also obtained from the TCGA database. Because the data were extracted from the TCGA database, following the publication guidelines strictly approved by TCGA, there was no requirement for ethics committee approval.

The association between smoking cessation and overall survival

The smoking status was included current smokers (included daily smokers and non-daily smokers or occasional smokers) and current reformed smokers (people who were not smoking at the time of the interview but have smoked at least 100 cigarettes in their life). To understand the association between smoking cessation and patients’ overall survival, age and multivariate-adjusted Cox regressions were performed to calculate the hazard ratio (HR) with 95% confidence intervals (CIs). In the multivariate model, we adjusted for age, gender, tumor stage, tumor status and HPV status.

Differentially expressed gene analysis

Differentially expressed mRNAs (DEGs), lncRNAs and miRNAs were identified between current smokers and reformed smokers (p-value <0.05) by “limma” package with R(15). Function analysis including Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway was performed by “clusterProfiler” package in R(16).

Somatic mutation analysis

The somatic mutation frequency >20 was considered to compare their relative distribution between current smokers and reformed smokers. Waterfall-map for somatic mutation patterns was performed by the R package “GenVisR”(17). Then the association between gene expression and somatic mutation was determined by the Mann-Whitney U test. The total mutation loads of current smokers and reformed smokers were compared using the Mann-Whitney U test.

Copy number variation analysis

Values of Segment Mean bigger than 0.2 were defined as gain and less than -0.2 as loss(18). Chi-test was used to compare CNV between current smokers and reformed smokers. Circos plots were performed by the R package “Rcircos”(19). Then the association between gene expression and CNV was determined by the Kruskal-Wallis test.

DNA Methylation analysis

The gene methylation matrix was normalized by “limma” package with R. The gene with different DNA methylation level between current smokers and reformed smokers were also used “limma” package. Then the association between gene expression and DNA methylation level was determined by the Pearson correlation coefficient. Above DEGs related to SNV, CNV or DNA methylation were considered as the key DEGs.
**Construction of competing endogenous RNAs network**

The differentially expressed lncRNAs, miRNAs and key DEGs were used to construct the competing endogenous RNAs (ceRNA) network. LncRNA-miR links were predicted by Mircode database (20). The targets of miRNAs were predicted by miRDB 6.0 (21), mirTarBase 7.0 (22), and TargetScan 7.2 (23). The ceRNA network was visualized by Cytoscape 3.6 (24).

**Immune cell scores**

The abundance of immune cells was measured using three different algorithms, including the Tumor IMMune Estimation Resource (TIMER, six immune cell types) (25), Cell type Identification by Estimating Relative Subsets of RNA Transcripts (CIBERSORT, 22 immune cell types) (26) and xCell (64 immune and stromal cell types) (26). The comparison of immune cell distribution between current smokers and reformed smokers using the Mann-Whitney U test.

**Construction and validation of a smoking signature**

To evaluate the true degree of smoking and predict the overall survival of smokers, we provided a quantitative smoking signature using key DEGs and immune cells. According to the somatic mutation, genes were valued as 0 (wild) and 1 (mutation). According to the copy number variation, genes were valued as -1 (loss), 0 (normal) and 1 (gain). The immune cell fraction level was valued as 0 or 1. A value of one was assigned when the fraction of one type of cell was over the corresponding cut-off value and a value of zero otherwise. Smoking status was also important and included in the smoking signature (current smoking=1, stopping smoking=0). First, prognostic factors were identified by performing univariate Cox regression analysis. LASSO-COX analysis was then employed to obtain the most useful predictive features. The smoking signature was built based on the corresponding coefficients.

The Kaplan-Meier (K-M) survival curves were performed to evaluate the distinguishing ability of patients’ outcome, and time-dependent receiver operating characteristic (survival ROC) curves were applied to assess the predicting power of smoking signature. To provide a quantitative tool for predicting the individual probability of patients’ outcomes, we built a prognostic nomogram on the basis of the smoking signature and clinical information in smokers with LUSC. Calibration curves for two-year, three-year and five-year were also plotted to compare the predicted and actual probabilities.

We calculated the smoking signature in different cancers in the TCGA database using the same formula. To confirm the applicability and reliability of the smoking signature, age and multivariate-adjusted Cox regressions were performed to calculate the hazard ratio (HR) with 95% confidence intervals (CIs). In the multivariate model, we adjusted for age, gender, tumor stage, tumor status and HPV status.

**Statistics**

All data were expressed as mean ± SD (standard deviation). LASSO-COX analysis was performed using the “glmnet” package. Survival ROC was plotted using the “survivalROC” package. The optimal cut-off values of each immune cell type were evaluated based on the patients’ overall survival and cell fraction using the “survminer” package. Survival analysis was used Cox proportional hazards model and survival distributions in different groups were visualized using Kaplan-Meier curves. The above analysis was conducted using R software 3.5 and SPSS software 23.0, and all statistical tests were two-sided and P < 0.05 was considered statistically significant.

**Results**

**Quitting smoking can significantly improve the prognosis of cancer patients**

The age and multivariable-adjusted hazard ratios for the association between smoking cessation and patients’ overall survival are presented in table 1. In age-adjusted models, quitting smoking was significantly associated
with longer survival time in CESC, HNSC, LUSC and PAAD. In the multivariable-adjusted model, though there was no significance in the majority of cancers, reformed smokers had a better prognosis than current smokers. Only in LUSC, the multivariable-adjusted HR (95% CI) was 0.67 (0.48-0.94) among reformed smokers relative to current smokers, indicating that quitting smoking was the independent protective factor for prognosis. Therefore, we further explored the potential mechanism by which quitting smoking can improve the prognosis in LUSC.

**Differentially expressed gene analysis**
Differentially expressed mRNA, lncRNA and miRNA between reformed smokers and current smokers. In total, 2,899 differentially expressed genes (DEGs) \( p<0.05 \) were identified, including 2,102 genes with significantly lower expression and 797 genes with higher expression in reformed smokers than current smokers (Figure 1A, Supplementary table 2). GO and KEGG analysis showed that DEGs mainly enriched in the DNA and RNA-related pathways and GO terms, including DNA replication, RNA splicing and others (Supplementary figure 1, Supplementary table 3 and 4). Similarly, a total of 48 differentially expressed miRNAs (20 down-regulated and 28 up-regulated miRNAs, Figure 1B) and 1,326 differentially expressed lncRNAs (1,207 down-regulated and 119 up-regulated miRNAs, Figure 1C)

**Differences in somatic mutations related to smoking status**
To reveal the relevant genetic alterations, we analyzed the somatic mutations between current smokers and reformed smokers. While no significant difference was found for total mutation load (Supplementary figure 2A), relative mutations frequency of 71 genes was significantly different (Figure 2A, Supplementary table 5) between reformed smokers and current smokers. Among them, there were ten DEGs (Supplementary Figure 2B). We assessed whether these DEGs transcriptions were affected by somatic mutations and found that the expression of GPATCH8 and ZFC3H1 was significantly associated with their somatic mutations (Figure 2B, Supplementary Figure 2C).

**Differences in copy number variations related to smoking status**
We found 781 genes with different CNV and their copy number gains or loss mainly on chromosome 19, 1 and 17 (Figure 2C, Supplementary table 6). Among them, we assessed whether copy number variations affected transcription of 94 DEGs (Supplementary figure 3A) and found that 73 DEGs expression was closely related to their copy number variations (Supplementary table 7).

**Differences in DNA methylation related to smoking status**
To explore the impact of smoking on DNA methylation, we analyzed the gene methylation levels. We found that 964 genes in reformed smokers with different DNA methylation compared to current smokers (Figure 2D, Supplementary table 8). Among 77 DEGs (Supplementary figure 3B), we were interested in ten DEGs whose expression were significantly associated with methylation level \( \text{Cor}<0.30, p<0.05 \), including HOXB2 \( (\text{Cor}=-0.728, p<0.001) \) and PTHLH \( (\text{Cor}=-0.565, p<0.001) \)(Figure 2E and Supplementary figure 3C).

In summary, the above analysis indicated that 85 key DEGs affected by different modes of genetic and epigenetic regulation might represent key drivers in smokers.

**Construction of ceRNAs network**
Next, we constructed the ceRNA network using differentially expressed lncRNAs, miRNAs and key DEGs. Target miRNA prediction revealed 139 lncRNA-miRNA links, including 76 lncRNAs and eight miRNAs according to the Mircode database (Supplementary table 9). Target gene prediction for above eight miRNAs revealed 3667 miRNA-mRNA links (prediction in at least two out of three databases) (Supplementary table 10). Based on lncRNA-miRNA and miRNA-mRNA links, a lncRNA-miRNA-DEGs complex network (69 lncRNAs, 5 miRNAs and 13 DEGs) was built to summarize underlying molecular traits of smokers (Figure 1D).

**Estimation of immune cell type fractions in LUSC**
We estimated the abundance of immune cells using three different algorithms. The distribution of several
immune cell fractions in reformed smoker was different from that in current smokers, including CD8+ T cell (TIMER), follicular helper T cell (CIBERSORT), gamma delta T cell (CIBERSORT), M0 Macrophage (CIBERSORT), central memory CD4+ T cell (XCELL) and central memory CD8+ T cell (XCELL) (Supplementary figure 4).

Construction and validation of smoking signature

Univariate Cox regression analysis and LASSO-COX analysis were performed to identify key prognostic markers, and smoking signature was built (Supplementary figure 5, Supplementary table 11). The formula for the smoking signature was based on the corresponding coefficients (Supplementary table 12): smoking signature = 0.5410* (smoking status) + 0.3278*ZFC3H1|snv +0.2153* GPATCH8|snv + 0.3625* NOL8|cnv + -0.5947* RPL10A|cnv + -0.3870* follicular helper T cell (CIBERSORT) + 0.5414* M0 Macrophage (CIBERSORT) + -0.1420* central memory CD8+ T cell (XCELL).

Distributions of the smoking signature in smokers were showed that reformed smokers had lower smoking signature than current smokers (Figure 3A). The Kaplan-Meier curves were plotted to confirm that the patients with high smoking signature had poorer prognosis (p<0.001, Figure 3B). The smoking signature also exhibited the strong predicted power of 2, 3 and 5-year survival (AUC= 0.65, 0.67 and 0.70, Figure 3C). Moreover, univariate and multivariate Cox regression analysis showed that smoking signature could become potential independent prognostic indicators (p<0.001) (Figure 3D-E). To provide a quantitative tool to predict patients’ survival, we constructed the prognostic nomogram integrating smoking signature and clinical information in smokers with LUSC (Figure 3F). Moreover, the calibration curve of the nomogram demonstrated good agreement between prediction and observation (Figure 3G).

To confirm the applicability and reliability of the smoking signature, we verified it in various cancers. The smoking signature of each type of cancer was provided in Supplementary table 13. In age-adjusted models, the smoking signature was significantly associated with overall survival in BLCA, CESC, HNSC, LUAD, LUSC and PAAD. In multivariable-adjusted model, patients with higher smoking signature had higher hazard rates than patients with lower smoking signature in BLCA (HR=1.70, 95% CI=1.01-2.88), CESC (HR=5.69, 95% CI=1.37-23.69), HNSC (HR=1.97, 95% CI=1.41-2.76), LUAD (HR=1.73, 95% CI=1.16-2.57), LUSC (HR=1.70, 95% CI=1.19-2.43) and PAAD (HR=4.28, 95% CI=1.47-12.47) (Table 2). We also constructed the prognostic nomogram with the smoking signature in BLCA, CESC, ESCA, HNSC, LUAD, LUSC and PAAD as quantitative tools (Supplementary figure 6).

Discussion

Tobacco smoking is an established risk factor for many cancers’ development. However, it is not clear how quitting smoking affects the prognosis of malignant tumors, and the lack of a special tobacco smoking assessment signature potentially underestimates the true impact of smoking on cancer survival. In the present study, we used the TCGA cohort to estimate the association between smoking cessation and overall survival, understand the genetic and immune microenvironment of smoking-patients, and provide an effective smoking signature for evaluating the smoking level to predict prognosis.

By performing age-adjusted Cox regressions, we found that smoking at diagnosis increased mortality risk as compared with reformed smokers in CESC, HNSC, LUSC and PAAD. Importantly, quitting smoking was the independent prognostic factor of LUSC. Several studies have evaluated the effect of smoking on LUSC. Molinier et al.(27) estimated five-year survival in non-small-cell lung cancer patients and found that smoking level at diagnosis was an independent negative prognostic factor in LUSC patients. Nakamura et al.(28) performed the multivariate analysis to find that smoking in LUSC was associated with recurrence-free survival. Although without distinguishing histological type, a cohort claimed that recent quitting could decrease the risk of death among patients with lung cancer (HR, 0.90; 95% CI 0.81–0.99)(11). Synthesizing above all outcomes, it is demonstrated reducing smoking could decrease the deterioration risk as compared with current smoking, suggesting a reversible effect of smoking in LUSC.
Then we analyzed the somatic mutation, copy number variation and DNA methylation between reformed smokers and current smokers. A total of 85 key DEGs were identified, whose expression was regulated by gene mutation or methylation. Among them, several key genes have been confirmed to be associated with lung squamous cell carcinoma. CBLC can be recruited into the epidermal growth factor receptor (EGFR) to increases EGFR ubiquitination, and thereby downregulate EGFR signaling in lung cancer patients(29). Zhan et al.(30) suggested that RPS11 was considered as the suitable reference gene for qRT-PCR-based studies of squamous cell lung carcinoma because of its high and stable expression. Sienel et al.(31) found that CEACAM1 has implicated in the development and progression of LUSC and an independent prognosticator for survival. Besides, these key genes also played important roles in smoking-related cancers, including lung cancer. Shui et al.(32) found the DNA methylation of LGALS3 was associated with smoking status in prostate cancer and strongly correlated with its expression. DNA Ligase I (LIG1) is a DNA repair gene involved in both the nucleotide excision repair (NER) and the base excision repair (BER) pathways(33). Many studies confirmed that variants in LIG1 may predispose to smoking-related lung cancer(34,35). By comparing the gene expression profiles in lung cancer between non-smokers and smokers, Woenckhaus et al.(36) found PTHLH, being involved in matrix degradation, was differentially expressed, which could reflect early cigarette smoke-induced and cancer-relevant molecular lesions. Chronic obstructive pulmonary disease (COPD) is another threat of smoking-induced lung injury, which can be the driving factor for lung cancer(37). ATG7, an autophagic gene, is increasingly activated in the early stages of lung injury induced by cigarette smoke(38,39). AXL is a receptor tyrosine kinase related to cancer and immune function, which mediates signal transduction related to proliferation and inflammation(40). During secondhand smoke, the interaction between AXL and receptors for advanced glycation end-products to cause COPD. Exposure to cigarette smoke, LGALS3 can increase CXCL8 secretion to induce inflammation(41) in COPD. Nowadays, the importance of smoking cessation in the management of COPD has been well-established(42). Similarly, cancer patients should quit smoking as soon as possible, which is helpful for cancer treatment by regulating key genes. In addition, we constructed the ceRNA network to summarize the underlying molecular traits of smokers, indicating that smoking could affect DEGs by different modes of genetic and epigenetic regulation.

Therefore, we provided a comprehensive smoking signature including immune microenvironment and epigenetic regulation to evaluate the true impact of smoking because of the complexity in cancer smokers. By understanding the immune microenvironment, we found that the fractions of follicular helper T cell, M0 Macrophage and central memory CD8+ T cell were different between reformed smokers and current smokers, suggesting that smoking status could change immune microenvironment to affect prognosis. It is reported that immune homeostasis in tumor microenvironment appears to be less compromised in non-smokers than in ever-smokers. In addition, the composition of leukocyte subtypes is closely correlated not only with smoking history, but also with patients’ outcome(43). In the immune response, different subsets of T cells play different roles, such as releasing of lymphokines, killing target cells, assisting immune response, memory-specific antigen stimulation(44). Follicular helper T cells play crucial roles in the development of humoral immunity(45). Yang et al.(13) estimated the immune cell type fractions in digestive system cancer and found that follicular helper T cells were the protective factors of patients’ overall survival. Many studies also have shown that CD8+ T cells usually mean a better prognosis among cancer patients(46,47). M0 macrophages were reported to be inversely associated with patients’ outcomes in various, such as adrenal cortical carcinoma and lung cancer(48,49). Nowadays, majority studies evaluated the degree of smoking based on the frequency of tobacco use to define heavy smokers and light smokers without uniform quantifying standards(50-52). Moreover, these cut-offs cannot accurately identify the true degree of smoking because they do not comprehensively consider the DNA damage and microenvironment alteration. Rosenthal et al.(53) developed deconstructSigs to identify mutational smoking signature in LUSC, LUAD and HNSC. Desrichard et al. confirmed that patients with high mutational smoking signature had poorer overall survival in HNSC (HR =1.50, 95% CI =1.23-1.81, P <0.01), but the mutational smoking signature was not prognostic in LUSC (HR =1.02, 95% CI =0.71-1.46, P =0.92) (54) and LUAD (HR =1.18, 95% CI =0.46-3.04, P =0.74). Importantly, the smoking signature we constructed not only can predict the overall survival in LUSC, but also can serve as prognostic indicators in BLCA, CESC, HNSC, LUAD and PAAD.

**Conclusion**
In the present study demonstrated that quitting smoking at diagnosis decreased the risk of death in cancer patients, suggesting a reversible effect of smoking on prognosis. We further provided the smoking signature by understanding the underlying molecular traits to evaluate the true effect of smoking, which could improve the prognostic prediction. At the same time, we suggested that smoking cessation could be a part of cancer treatment to improve the survival rate of cancer patients.

### Abbreviations

| Abbreviation | Description |
|--------------|-------------|
| TCGA         | the Cancer Genome Atlas |
| HR           | Hazard ratio |
| CI           | confidence interval |
| LUSC         | lung squamous cell carcinoma |
| DEG          | differential expression gene |
| BLCA         | bladder urothelial carcinoma |
| CESC         | cervical squamous cell carcinoma and endocervical adenocarcinoma |
| ESCA         | esophageal carcinoma |
| HNSC         | head and neck squamous cell carcinoma |
| LUAD         | lung adenocarcinoma |
| PAAD         | pancreatic adenocarcinoma |
| CNV          | copy number variation |
| GO           | Gene Ontology |
| KEGG         | Kyoto Encyclopedia of Genes and Genomes |
| ceRNA        | competing endogenous RNAs |
| TIMER        | Tumor IMmune Estimation Resource |
| CIBERSORT    | Cell type Identification by Estimating Relative Subsets of RNA Transcripts |
| ROC          | receiver operating characteristic |
| COPD         | chronic obstructive pulmonary disease |

### Declarations
Ethics approval and consent to participate

Because the data were extracted from the TCGA database, following the publication guidelines strictly approved by TCGA, there was no requirement for ethics committee approval.

Consent for publication

Not applicable.

Availability of data and materials

The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request. The data that support the findings of this study were derived from the following resource: TCGA database, https://www.cancer.gov/tcga.

Competing interests

The authors declare no conflict of interest.

Funding

The present study was supported by the National Natural Science Foundation of China (81673132, 81972998), and Postgraduate Research & Practice Innovation Program of Jiangsu Province (KYCX19_0120).

Authors' contributions

Yang Sheng performed the research and wrote the manuscript. Liu Tong and Liang Geyu revised the manuscript. All authors read and approved the final manuscript.

Acknowledgements

The present study was supported by the National Natural Science Foundation of China (81673132, 81972998), and Postgraduate Research & Practice Innovation Program of Jiangsu Province (KYCX19_0120).

References

1. Kispert S, Marentette J, McHowat J. Cigarette smoking promotes bladder cancer via increased platelet-activating factor. Physiological reports 2019;7(3):e13981 doi 10.14814/phy2.13981.
2. Jethwa AR, Khariwala SS. Tobacco-related carcinogenesis in head and neck cancer. Cancer metastasis reviews 2017;36(3):411-23 doi 10.1007/s10555-017-9689-6.
3. Schwartz AG, Cote ML. Epidemiology of Lung Cancer. Advances in experimental medicine and biology 2016;893:21-41 doi 10.1007/978-3-319-24223-1_2.
4. Ilic M, Ilic I. Epidemiology of pancreatic cancer. World journal of gastroenterology 2016;22(44):9694-705 doi 10.3748/wjg.v22.i44.9694.
5. Alexandrov LB, Ju YS, Haase K, Van Loo P, Martincorena I, Nik-Zainal S, et al. Mutational signatures associated with tobacco smoking in human cancer. Science 2016;354(6312):618-22 doi 10.1126/science.aag0299.
6. Sopori M. Effects of cigarette smoke on the immune system. Nature reviews Immunology 2002;2(5):372-7 doi 10.1038/nri803.
7. Du E, Mazul AL, Farquhar D, Brennan P, Anantharaman D, Abedi-Ardekani B, et al. Long-term Survival in Head and Neck Cancer: Impact of Site, Stage, Smoking, and Human Papillomavirus Status. The Laryngoscope 2019;129(11):2506-13 doi 10.1002/lary.27807.
8. Yuan C, Morales-Oyarvide V, Babic A, Clish CB, Kraft P, Bao Y, et al. Cigarette Smoking and Pancreatic Cancer Survival. Journal of clinical oncology : official journal of the American Society of Clinical Oncology 2017;35(16):1822-8 doi 10.1200/JCO.2016.71.2026.
9. Warren GW, Kasza KA, Reid ME, Cummings KM, Marshall JR. Smoking at diagnosis and survival in cancer patients. International journal of cancer 2013;132(2):401-10 doi 10.1002/ijc.27617.
10. Tao L, Wang R, Gao YT, Yuan JM. Impact of postdiagnosis smoking on long-term survival of cancer patients: the Shanghai cohort study. Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology 2013;22(12):2404-11 doi 10.1158/1055-9965.EPI-13-0805-T.
11. Tabuchi T, Goto A, Ito Y, Fukui K, Miyashiro I, Shinozaki T. Smoking at the time of diagnosis and mortality in cancer patients: What benefit does the quitter gain? International journal of cancer 2017;140(8):1789-95 doi 10.1002/ijc.30601.
12. Psyrri A, DiMaio D. Human papillomavirus in cervical and head-and-neck cancer. Nature clinical practice Oncology 2008;5(1):24-31 doi 10.1038/ncompcancer0984.
13. Yang S, Liu T, Cheng Y, Bai Y, Liang G. Immune cell infiltration as a biomarker for the diagnosis and prognosis of digestive system cancer. Cancer science 2019;110(12):3639-49 doi 10.1111/cas.14216.
14. Song J, Xu Q, Zhang H, Yin X, Zhu C, Zhao K, et al. Five key IncRNAs considered as prognostic targets for predicting pancreatic ductal adenocarcinoma. Journal of cellular biochemistry 2018;119(6):4559-69 doi 10.1002/jcb.26598.
15. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic acids research 2015;43(7):e47 doi 10.1093/nar/gkv007.
16. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. Omics : a journal of integrative biology 2012;16(5):284-7 doi 10.1089/omi.2011.0118.
17. Skidmore ZL, Wagner AH, Lesurf R, Campbell KM, Kunisaki J, Griffith OL, et al. GenVisR: Genomic Visualizations in R. Bioinformatics 2016;32(19):3012-4 doi 10.1093/bioinformatics/btw325.
18. Feng B, Shen Y, Pastor X, Bieg M, Plath M, Ishaque N, et al. Integrative analysis of multi-omics data identified EGFR and PTGS2 as key nodes in a gene regulatory network related to immune phenotypes in head and neck cancer. Clinical cancer research : an official journal of the American Association for Cancer Research 2020 doi 10.1158/1078-0432.CCR-19-3997.
19. Zhang H, Meltzer P, Davis S. RCircos: an R package for Circos 2D track plots. BMC bioinformatics 2013;14:244 doi 10.1186/1471-2105-14-244.
20. Jeggari A, Marks DS, Larsson E. miRcode: a map of putative microRNA target sites in the long non-coding transcriptome. Bioinformatics 2012;28(15):2062-3 doi 10.1093/bioinformatics/bts344.
21. Wong N, Wang X. miRDB: an online resource for microRNA target prediction and functional annotations. Nucleic acids research 2015;43(Database issue):D146-52 doi 10.1093/nar/gku1104.
22. Chou CH, Shrestha S, Yang CD, Chang NW, Lin YL, Liao KW, et al. miRTarBase update 2018: a resource for experimentally validated microRNA-target interactions. Nucleic acids research 2018;46(D1):D296-D302 doi 10.1093/nar/gkx1067.
23. Grimson A, Farh KK, Johnston WK, Garrett-Engele P, Lim LP, Bartel DP. MicroRNA targeting specificity in mammals: determinants beyond seed pairing. Molecular cell 2007;27(1):91-105 doi 10.1016/j.molcel.2007.06.017.
24. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome research 2003;13(11):2498-504 doi 10.1101/gr.1239303.
25. Li T, Fan J, Wang B, Traugh N, Chen Q, Liu JS, et al. TIMER: A Web Server for Comprehensive Analysis of Tumor-Infiltrating Immune Cells. Cancer research 2017;77(21):e108-e10 doi 10.1158/0008-5472.CAN-17-0307.
26. Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, et al. Robust enumeration of cell subsets from tissue expression profiles. Nature methods 2015;12(5):453-7 doi 10.1038/nmeth.3337.
27. Molinier O, Goupil F, Debieuvre D, Auliac JB, Jeandeau S, Lacroix S, et al. Five-year survival and prognostic factors according to histology in 6101 non-small-cell lung cancer patients. Respiratory medicine and research 2019;77:46-54 doi 10.1016/j.resmer.2019.10.001.
28. Nakamura H, Sakai H, Kimura H, Miyazawa T, Marushima H, Saji H. Difference in Posttreatment Prognostic Factors between Lung Adenocarcinoma and Squamous Cell Carcinoma. Annals of thoracic and cardiovascular surgery : official journal of the Association of Thoracic and Cardiovascular Surgeons of Asia 2017;23(6):291-7 doi 10.5761/atcs oa.17-00020.
29. Kim B, Lee HJ, Choi HY, Shin Y, Nam S, Seo G, et al. Clinical validity of the lung cancer biomarkers identified by bioinformatics analysis of public expression data. Cancer research 2007;67(15):7431-8 doi 10.1158/0008-5472.CAN-07-0003.
30. Zhan C, Zhang Y, Ma J, Wang L, Jiang W, Shi Y, et al. Identification of reference genes for qRT-PCR in human lung squamous-cell carcinoma by RNA-Seq. Acta biochimica et biophysica Sinica 2014;46(4):330-7 doi 10.1093/abbs/gmt153.
31. Sienel W, Dango S, Woelfle U, Morresi-Hauf A, Wagener C, Brummer J, et al. Elevated expression of carcinomaembryonic antigen-related cell adhesion molecule 1 promotes progression of non-small cell lung cancer. Clinical cancer research : an official journal of the American Association for Cancer Research 2003;9(6):2260-6.
32. Shui IM, Wong CJ, Zhao S, Kolb S, Ebot EM, Geybels MS, et al. Prostate tumor DNA methylation is associated with cigarette smoking and adverse prostate cancer outcomes. Cancer 2016;122(14):2168-77 doi 10.1002/cncr.30045.
33. Lee YC, Morgenstern H, Greenland S, Tashkin DP, Papp J, Sinsheimer J, et al. A case-control study of the association of the polymorphisms and haplotypes of DNA ligase I with lung and upper-aerodigestive-tract cancers. International journal of cancer 2008;122(7):1630-8 doi 10.1002/ijc.23274.
34. Leng S, Stidley CA, Liu Y, Edlund CK, Willink RP, Han Y, et al. Genetic determinants for promoter hypermethylation in the lungs of smokers: a candidate gene-based study. Cancer research 2012;72(3):707-15 doi 10.1158/0008-5472.CAN-11-3194.
35. Sakoda LC, Loomis MM, Doherty JA, Julianto L, Barnett MJ, Neuhouser ML, et al. Germ line variation in nucleotide excision repair genes and lung cancer risk in smokers. International journal of molecular epidemiology and genetics 2012;3(1):1-17.
36. Woenckhaus M, Klein-Hitpass L, Grepmeier U, Merk J, Pfeiffer M, Wild P, et al. Smoking and cancer-related gene expression in bronchial epithelium and non-small-cell lung cancers. The Journal of pathology 2006;210(2):192-204 doi 10.1002/path.2039.
37. Eapen MS, Hansbro PM, Larsson-Callerfelt AK, Jolly MK, Myers S, Sharma P, et al. Chronic Obstructive Pulmonary Disease and Lung Cancer: Underlying Pathophysiology and New Therapeutic Modalities. Drugs 2018;78(16):1717-40 doi 10.1007/s40265-018-1001-8.
38. Fujita Y, Araya J, Ito S, Kobayashi K, Kosaka N, Yoshioka Y, et al. Suppression of autophagy by extracellular vesicles promotes myofibroblast differentiation in COPD pathogenesis. Journal of extracellular vesicles 2015;4:28388 doi 10.3402/jev.v4.28388.
39. Chen ZH, Kim HP, Sciufrba FC, Lee SJ, Feghali-Bostwick C, Stolz DB, et al. Egr-1 regulates autophagy in cigarette smoke-induced chronic obstructive pulmonary disease. PLoS one 2008;3(10):e3316 doi 10.1371/journal.pone.0003316.
40. Tsai KYF, Hirschi Budge KM, Llavina S, Davis T, Long M, Bennett A, et al. RAGE and AXL expression following secondhand smoke (SHS) exposure in mice. Experimental lung research 2019;45(9-10):297-309 doi 10.1080/01902148.2019.1684596.
41. Pouwels SD, Hesse L, Faiz A, Lubbers J, Bodha PK, Ten Hacken NH, et al. Susceptibility for cigarette smoke-induced DAMP release and DAMP-induced inflammation in COPD. American journal of physiology Lung cellular and molecular physiology 2016;311(5):L881-L92 doi 10.1152/ajplung.00135.2016.
42. Peiffer G, Undern M, Perriot J. [COPD and smoking cessation: Patients' expectations and responses of health professionals]. Revue de pneumologie clinique 2018;74(6):375-90 doi 10.1016/j.pneumo.2018.10.001.
43. Li X, Li J, Wu P, Zhou L, Lu B, Ying K, et al. Smoker and non-smoker lung adenocarcinoma is characterized by distinct tumor immune microenvironments. Oncoimmunology 2018;7(10):e1494677 doi 10.1080/2162402X.2018.1494677.
44. Speiser DE, Ho PC, Verdeil G. Regulatory circuits of T cell function in cancer. Nature reviews Immunology 2016;16(10):599-611 doi 10.1038/nri.2016.80.
45. Ng KW, Marshall EA, Enfield KS, Martin SD, Milne K, Pewarchuk ME, et al. Somatic mutation-associated T follicular helper cell elevation in lung adenocarcinoma. Oncoimmunology 2018;7(12):e1504728 doi 10.1080/2162402X.2018.1504728.
46. Fridman WH, Pages F, Sautes-Fridman C, Galon J. The immune contexture in human tumours: impact on clinical outcome. Nature reviews Cancer 2012;12(4):298-306 doi 10.1038/nrc3245.
47. Kawai O, Ishii G, Kubota K, Murata Y, Naito Y, Mizuno T, et al. Predominant infiltration of macrophages
and CD8(+) T Cells in cancer nests is a significant predictor of survival in stage IV nonsmall cell lung cancer. Cancer 2008;113(6):1387-95 doi 10.1002/cncr.23712.

48. Huang R, Liu Z, Tian T, Song D, Yan P, Yin H, et al. The construction and analysis of tumor-infiltrating immune cells and ceRNA networks in metastatic adrenal cortical carcinoma. Bioscience reports 2020;40(3) doi 10.1042/BSR20200049.

49. Liu X, Wu S, Yang Y, Zhao M, Zhu G, Hou Z. The prognostic landscape of tumor-infiltrating immune cell and immunomodulators in lung cancer. Biomedicine & pharmacotherapy = Biomedecine & pharmacotherapie 2017;95:55-61 doi 10.1016/j.biopha.2017.08.003.

50. Chin DL, Hong O, Gillen M, Bates MN, Okechukwu CA. Heavy and light/moderate smoking among building trades construction workers. Public health nursing 2013;30(2):128-39 doi 10.1111/j.1525-1446.2012.01041.x.

51. Perez-Rubio G, Garcia-Carmona S, Garcia-Gomez L, Hernandez-Perez A, Ramirez-Venegas A, Lopez-Flores LA, et al. The VNTR 48 bp Polymorphism in the DRD4 Gene Is Associated with Higher Tobacco Smoking in Male Mexican Mestizo Smokers with and without COPD. Diagnostics 2019;10(1) doi 10.3390/diagnostics10010016.

52. Kim SK, Kim HC, Shim JS, Kim DJ. Effects of cigarette smoking on blood lipids in Korean men: Cardiovascular and Metabolic Diseases Etiology Research Center cohort. The Korean journal of internal medicine 2020;35(2):369-82 doi 10.3904/kjim.2019.133.

53. Rosenthal R, McGranahan N, Herrero J, Taylor BS, Swanton C. DeconstructSigs: delineating mutational processes in single tumors distinguishes DNA repair deficiencies and patterns of carcinoma evolution. Genome biology 2016;17:31 doi 10.1186/s13059-016-0893-4.

54. Desrichard A, Kuo F, Chowell D, Lee KW, Riaz N, Wong RJ, et al. Tobacco Smoking-Associated Alterations in the Immune Microenvironment of Squamous Cell Carcinomas. Journal of the National Cancer Institute 2018;110(12):1386-92 doi 10.1093/jnci/djy060.

### Tables

| Cancer type | Current smoker | Reformed smoker | Age-adjusted | P value | MV-adjusted a | P value |
|-------------|----------------|-----------------|--------------|---------|---------------|---------|
| BLCA        | 91             | 198             | 0.79(0.54-1.16) | 0.230   | 0.72(0.42-1.22) | 0.219   |
| CESC b      | 64             | 53              | **0.34(0.16-0.71)** | **0.004** | 0.63(0.26-1.53) | 0.309   |
| ESCA        | 37             | 73              | 0.89(0.46-1.72) | 0.723   | 0.89(0.39-2.02) | 0.780   |
| HNSC b      | 178            | 215             | **0.65(0.47-0.89)** | **0.007** | 0.67(0.44-1.02) | 0.059   |
| LUAD        | 122            | 311             | 1.12(0.77-1.63) | 0.561   | 1.34(0.85-2.10) | 0.207   |
| LUSC        | 134            | 340             | **0.62(0.46-0.83)** | **0.001** | **0.67(0.48-0.94)** | **0.020** |
| PAAD        | 20             | 60              | **0.45(0.23-0.89)** | **0.021** | 0.54(0.24-1.20) | 0.132   |

a: MV-adjusted for age (continuous), gender (female, male), tumor stage (stage I, stage II, stage III and stage IV) and tumor status (with tumor and tumor free).

b: MV-adjusted for age (continuous), gender (female, male), tumor stage (stage I, stage II, stage III and stage IV),
tumor status (with tumor and tumor free) and HPV status (positive and negative).

Table 2. the association between the smoking signature and patients’ overall survival

| Cancer type | Age-adjusted | P value | MV-adjusted | P value |
|-------------|--------------|---------|-------------|---------|
| BLCA        | 1.03(1.01-1.05) | 0.001   | 1.70(1.01-2.88) | 0.048   |
| CESC b      | 4.69(1.80-12.23) | 0.002   | 5.69(1.37-23.69) | 0.017   |
| ESCA        | 1.22(0.50-3.01) | 0.663   | 1.30(0.50-3.41) | 0.596   |
| HNSC b      | 2.14(1.55-2.96) | <0.001  | 1.97(1.41-2.76) | <0.001  |
| LUAD        | 1.52(1.07-2.17) | 0.020   | 1.73(1.16-2.57) | 0.007   |
| LUSC        | 1.85(1.29-2.64) | 0.001   | 1.70(1.19-2.43) | 0.003   |
| PAAD        | 5.90(2.60-13.41) | <0.001  | 4.28(1.47-12.47) | 0.008   |

a: MV-adjusted for age (continuous), gender (female, male), tumor stage (stage I, stage II, stage III and stage IV) and tumor status (with tumor and tumor free).

b: MV-adjusted for age (continuous), gender (female, male), tumor stage (stage I, stage II, stage III and stage IV), tumor status (with tumor and tumor free) and HPV status (positive and negative).
D

ETV5-AS1  AP001619.1  AL137145.1  AC006305.1  MIR210HG
RHOA-IT1  DLEU2L  AC010536.2  SHANK3  TRAPPC12-AS1  BCI
AC008737.1  TTC3-AS1  SPAG5-AS1  LINC00313  FTX
AC018690.1  AP001157.1  AP001029.2  AF235103.1  MAGI1-IT1
TET2-AS1  AC125494.1  ST7-OT4  LINC00086  SNHG1
LINC00472  AL162430.2  CRYM-AS1  AC008781.1  AC010536
LINC00470  EAF1-AS1  HNRNPU-AS1  ERVK13-1  LINC002
Figure 1

Differentially expressed genes, miRNAs and lncRNAs between reformed smokers and current smokers. A: the differentially expressed genes. B: differentially expressed genes miRNAs. C: differentially expressed lncRNAs. D: The network summarizes complex connections between differentially expressed lncRNAs (red), lncRNAs targeted miRNAs (green), and DEGs (yellow).
Figure 2
Differences in the mutational landscape and DNA methylation between reformed smokers and current smokers. 
A: Top 20 differentially mutated genes between reformed smokers and current smokers. B: the expression of GPATCH8 and ZFC3H1 was significantly associated with their somatic mutations. C: genes with different CNV and their copy number gains or loss mainly on chromosome 19, 1 and 17. D: genes in reformed smoker with different DNA methylation compared to current smokers. E: expression of HOXB2 and PTHLH was significantly associated with methylation level.
**D**

| Smoking signature | p-value | Hazard ratio |
|-------------------|---------|--------------|
| smoking_signature | <0.001  | 2.955(2.054-4.251) |
| age               | 0.008   | 1.024(1.006-1.042) |
| gender            | 0.964   | 1.009(0.684-1.488) |
| stage             | 0.002   | 1.392(1.133-1.710) |
| tumor_status      | <0.001  | 3.055(2.175-4.290) |

**F**

Points

- smoking_signature
Figure 3

The smoking signature of evaluating the true affect of smoking on overall survival. A: the distribution of smoking signature between reformed smokers and current smokers. B: Kaplan–Meier (KM) curves for patients with high-smoking signature and low-smoking signature. C: Survival receiver–operating characteristic (ROC) curves for 2, 3, 5-year prediction. D-E: the association between smoking signature and death risk using univariate (D) and multivariate (E) Cox regression analysis. F: nomogram with smoking signature for predicting 2-, 3-, and 5-year death risk. F: calibration curves of nomograms in terms of the agreement between predicted and actual 2-, 3-, and 5-year outcomes.

**Supplementary Files**

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

- [Supplementarytable213.xlsx](#)
- [Supplementarytable1.xlsx](#)
- [Supplementaryfigure.pdf](#)