SMOX expression predicts the prognosis of non-small cell lung cancer

Zhanghao Huang[1,2,3#], Shuo Wang[1,2,3#], Hai-Jian Zhang[4#], You Lang Zhou[4], Jia-Hai Shi[1,2]

[1]Nantong Key Laboratory of Translational Medicine in Cardiothoracic Diseases, and Research Institution of Translational Medicine in Cardiothoracic Diseases, Affiliated Hospital of Nantong University, Nantong, China; [2]Department of Thoracic Surgery, Affiliated Hospital of Nantong University, Nantong, China; [3]Medical College of Nantong University, Nantong, China; [4]Research Center of Clinical Medicine, Affiliated Hospital of Nantong University, Nantong, China

Contributions: (I) Conception and design: Z Huang; (II) Administrative support: JH Shi, YL Zhou; (III) Provision of study materials or patients: S Wang, HJ Zhang; (IV) Collection and assembly of data: Z Huang; (V) Data analysis and interpretation: Z Huang; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

[#]These authors contributed equally to this work.

Correspondence to: You Lang Zhou; Jia-Hai Shi. Affiliated Hospital of Nantong University, Nantong, China.
Email: zhouyoulang@ntu.edu.cn; sjh@ntu.edu.cn.

Background: The development of non-small cell lung cancer (NSCLC) is very rapid, and the effect of its treatment is often closely related to the diagnosis time of the disease. Therefore, simple and convenient tumor biomarkers are helpful for the timely diagnosis and prevention of NSCLC.

Methods: Through univariate and multivariate Cox regression analyses, SMOX was determined as an independent prognostic factor of GSE42127, GSE41271, GSE68465, and TCGA datasets. Furthermore, western blot, reverse transcription-polymerase chain reaction (RT-PCR), and immunohistochemical analysis were performed to confirm the predictive efficiency of SMOX expression in NSCLC.

Results: Patients were divided into high and low expression groups according to the median value of SMOX expression, and Kaplan-Meier curves of multiple datasets indicated that patients with low SMOX expression had a better survival rate. According to the analysis of immune infiltration, the immune microenvironment, and immune checkpoints, SMOX expression of the high and low groups showed differences in immunity in NSCLC. By comparing cancer and adjacent tissues using western blot analysis, RT-PCR and immunohistochemical analysis, we found that SMOX was highly expressed in tumor tissues and had low expression in adjacent tissues. Simultaneously, the Kaplan-Meier curve suggested that among the 155 NSCLC patients, those with low SMOX expression had better survival.

Conclusions: SMOX can be used as an effective predictive target for NSCLC.

Keywords: Biomarker; immunity; gene set enrichment analysis (GSEA); prognosis

Introduction

Non-small cell lung cancer (NSCLC) accounts for a large proportion of lung cancer. Because the diagnosis is late, the time window for surgical treatment is lost. There are several treatment options for advanced NSCLC, including targeted therapy, chemoradiotherapy, immunotherapy, chemoimmunotherapy, and radiofrequency ablation. However, the prognosis of advanced NSCLC is not ideal due to local and distant metastasis. Therefore, finding a therapeutic target for NSCLC can help realize the diagnosis and treatment of NSCLC at an early stage, thereby greatly improving the prognosis of NSCLC (1).

Page 2 of 17

Huang et al. SMOX as a prognostic biomarker of NSCLC

SMOX, which metabolizes polyamine spermine to spermidine and hydrogen peroxide (HO), is associated with an increased risk of gastric cancer in humans (2). There is evidence that reactive oxygen species (ROS) mediate tissue homeostasis, cell signaling, differentiation, and survival, and ROS are the basis of angiogenesis and tumor growth. In cancer, angiogenesis is a key phenomenon in developing new blood vessels to promote the supply of oxygen and nutrients for tumor progression. Abnormally increased ROS is closely related to the development and metastasis of cancer. Since HO is the main component of ROS, molecular imaging and selective treatment of high HO levels are very attractive for cancer treatment.

Furthermore, SMOX is a highly inducible enzyme whose dysregulation alters the homeostasis of polyamines. Oxidation products of polyamine metabolism can induce cell damage and death. This is also one of the reasons why we chose SMOX for research.

There is a complex and inevitable connection between tumors and immunity (3). At present, immune responses and immune escape have become more widely recognized mechanisms in tumors. Immune responses mainly include humoral immunity and cellular immunity, and cellular immunity is the main tumor immune response. As a specific immune response, it mainly produces an immune response to solid tumor cells with strong antigenicity, and humoral immunity plays a synergistic effect. As a non-specific immune response, circulating antibodies and the like mainly produce immune responses against tumor cells that are weakly antigenic and in a free state.

In contrast, the mechanism of tumor immune escape involves immune selection, antigen modulation, tumor-induced immune suppression, and blocking factors. Immune selection means that tumor cells with strong antigenicity can cause an effective immune response and be eliminated, and tumor cells with relatively weak antigenicity can escape the immune system's surveillance and proliferate. Antigen modulation refers to the reduction or loss of antigenicity on the surface of tumor cells to avoid the recognition and attack of the host immune system. Tumor-induced immunosuppression means that when tumor cells first appear, the amount of antigens is small and weak, and continuous stimulation of the host can induce immune tolerance and make the host lose the ability to respond to the tumor. Finally, blocking factors refer to factors that cause immune promotion, including blocking antibodies, excessive free antigens, and antigen-antibody complexes. Immune responses have been considered major

components of tumor elimination in radiotherapy, and radiation-induced tumor injury promotes radiotherapy vaccination, enhances cytotoxic immune responses, and enhances immunotherapy. The efficacy of immunotherapy depends on the degree of infiltration of immune cells in the immune microenvironment, a complex ecosystem formed by unique and interacting cell populations. Its composition is related to the prognosis of cancer and the response to clinical treatment. The tumor microenvironment (TME) interacts with tumor cells through various cytokines and cytokine receptors to regulate their stemness. The main obstacle of immunotherapy is that effector immune cells cannot fully penetrate the TME. Therefore, exploring the mechanism of the TME of NSCLC is beneficial to improve immunotherapy further. The discovery of prognostic markers is helpful for early diagnosis and treatment, and the combination of the 2 greatly improves the predictive performance for NSCLC. We present the following article following the REMARK reporting checklist (available at http://dx.doi.org/10.21037/atm-21-998).

Methods

Source and collation of data

The data involved in our research came from The Cancer Genome Atlas (TCGA) and Gene Expression Omnibus (GEO) databases, including genetic transcriptome analysis and clinical information. The survival prognosis information of patients was an important basis for our selection of TCGA and GEO datasets. Among them, GSE41271 contained 275 tumor samples, GSE42127 contained 176 tumor samples, and GSE68465 contained 443 samples. The TCGA datasets contained 999 tumor samples. Downloaded transcriptome data were standardized and corrected before analysis, and samples with incomplete information were deleted when the clinical information was integrated.

Establishment and verification of independent prognostic factors

Univariate and multivariate Cox regression analyses were used to calculate 95% confidence interval (CI) and hazard ratio (HR), aiming to determine independent prognostic factors for lung cancer. Univariate survival analysis was used to analyze clinical factors and survival rates, while multivariate survival analysis comprehensively evaluated the

relationship between multiple clinical factors and overall survival rates. P value <0.05 was generally considered to be statistically significant. The significance of the Kaplan-Meier survival curve was to show the relationship between the expression of SMOX and the survival of lung cancer. We divided various clinical factors into different subgroups, showed the effect size of the test factors in each subgroup through forest plots, and even visually displayed the difference between each subgroup (4).

Functional enrichment analysis

We performed gene set enrichment analysis (GSEA) on RNA-seq data in the TCGA database, which aimed to further understand the biological function of SMOX by analyzing Gene Ontology (GO) terms and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways (5,6). The number of permutations was set to 1000. Other parameters were set based on their default values. The enrichment pathways in each phenotype were classified based on the nominal P value and the normalized enrichment score (NES). By extending the GSEA, singe sample GSEA (SSGSEA) allows the definition of an enrichment fraction that represents the absolute enrichment of a gene set in each sample within a given dataset.

Immune infiltration analysis

CIBERSORT is a method used to characterize the cell composition of complex tissues from the gene expression profile. When applied to hematopoietic subpopulations in RNA mixtures extracted from fresh, frozen, and fixed tissues (including solid tumors), CIBERSORT outperforms other methods in terms of noise, unknown mixture content, and closely related cell types. CIBERSORT can perform large-scale analysis of RNA mixtures used for cell biomarkers and therapeutic targets (http://cibersort.stanford.edu). CIBERSORT is also a deconvolution algorithm based on gene expression that can be used to evaluate samples. The variation among the 22 tumor-infiltrating immune cells reflects the correlation. Through Monte Carlo sampling, CIBERSORT approximately sets the p value for deconvolution to determine the confidence of the result. Monte Carlo sampling is a random approximate inference method. In essence, it transforms the longitudinal distribution of the sample in space into the equivalent distribution of the sample in the horizontal direction. The screening condition for further analysis was CIBERSORT

P<0.05. Tumor mutational burden (TMB) refers to the total number of mutations per megabase in tumor tissues, the mutation density of tumor genes, including the total number of gene coding errors, and splicing and gene insertion or deletion errors. The higher the TMB, the higher the characteristic of the tumor, the more obvious the difference from normal tissues, and the more effective the immunotherapy (7).

Western blotting analysis

Proteins were extracted from tissues using 1× radioimmunoprecipitation. The lysis buffer was assayed with protease inhibitors. The samples were loaded on a 10% SDS polyacrylamide gel for electrophoresis and transferred to polyvinylidene fluoride film. Protein bands were identified by visualized primary antibodies using the Odyssey Infrared Imaging System (LI-COR Biological Sciences), and images were used for optical density analysis (8).

Reverse transcription-polymerase chain reaction (RT-PCR)

Total RNAs were extracted from the patients' tissues according to the instructions of the product specification. Reverse transcription of 1 μg of total RNA was performed using the M-MLV reverse transcriptase product (Promega). The ABI7500 real-time PCR system (Thermo Fisher Science) was used to perform RT-qPCR. Glyceraldehyde 3-phosphate dehydrogenase (GAPDH) was used as the internal control, and the level of SMOX was detected by RT-qPCR as follows: denatured at 94 ℃ for 5 s, annealed, and extended at 62 ℃ for 40 s for PCR amplification. The 1 Ct method was used to calculate the relative expression level of SMOX. Simply put, the difference between the GAPDH Ct value and the SMOX Ct value was defined as the 1Ct value. A high 1Ct value was considered to indicate a relatively low expression of SMOX in each sample. All PCR amplifications were performed in triplicate, and 3 independent experiments were repeated (9). SMOX forward primer was cag tgg tga cag tgc gga tga c. SMOX reverse primer was acc ctg ctc aag tgc ttt gg.

Immunohistochemical analysis

Lung cancer tissue sections were provided by the Affiliated Hospital of Nantong University. After lung cancer tissue was deparaffinized, hydrated, and blocked, SMOX

Page 4 of 17

Huang et al. SMOX as a prognostic biomarker of NSCLC

antibody (1:1,000, Proteintech) was added dropwise and incubated overnight at 4 ℃. Sections were then washed with phosphate-buffered saline (PBS) 3 times and then incubated with horseradish peroxidase (HRP) goat anti-rabbit IgG polyclonal antibody at room temperature for 30 minutes. DAB and hematoxylin were used to stain tissue slides. SMOX expression in lung cancer tissues was estimated by software (Image-Pro Plus 6.0) and evaluated by 2 independent investigators who did not know clinical information, and consensus resolved the disagreements. When the results of the two investigators diverged, a third investigator was invited to analyze (10).

Statistical analysis

One-way ANOVA and the Kruskal-Wallis test, non-parametric and parametric methods, were used to compare 2 or more groups. Spearman and distance correlation analysis was used to calculate the correlation coefficient. To identify genes in the analysis of differential genes, the Benjamini-Hochberg method converted the P value to false discovery rate (FDR). P values were two-sided, and P values of less than 0.05 were regarded as statistically significant.

Ethics declaration

The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013). The study was approved by ethics board of Affiliated Hospital of Nantong University (No.: 2020-L002).

Results

Establishment of independent prognostic factors in the GEO Database

We obtained genome-wide expression from multiple datasets (GSE41271, GSE42127, and GSE68465) in the GEO database, aiming to reflect the function of mRNAs in lung cancer. After annotating the expression data and sorting out the clinical data, the dataset could be subjected to survival analysis, independent prognosis, and clinical relevance filtering. After a series of screenings, GSE41271 screened out 466 independent prognostic factors, GSE42127 screened out 300 independent prognostic factors, while GSE68465 screened out 341 independent prognostic factors. A total of 16 common genes were obtained by taking the intersection of the

3 through a Venn diagram (*Figure 1A*, *Table 1*). The correlation of these 16 prognostic genes in the TCGA database was shown through a network diagram, and the selected correlation threshold was 0.3 (*Figure 1B*). Since the functional properties of most of these genes have been thoroughly explored, while the functional properties of SMOX needed further exploration, we chose the SMOX gene and explored its functional role in lung cancer. The survival analysis curves of SMOX in the 3 datasets are shown in *Figure 1C*. Simultaneously, univariate and multivariate Cox regression analyses found that SMOX was indeed an independent prognostic factor for lung cancer and could be displayed in forest plots (*Figure 1D,E*).

Clinical correlation analysis

The TNM staging of lung cancer in the clinical staging of the location, size, and spread of lung cancer. T contains T1-T4, N contains N1-N3, and M contains M0-M1. All 3 datasets contained stage information, but GSE42127 and GSE685465 only contained stages I-III, as there were no patients with stage IV. It is worth noting that only patients in GSE68465 had clinical information of T and N. Since no patients had N3 disease, they were classified as N0-N2. Clinical correlation analysis displayed that in the GSE41271 dataset, there were significant differences in the expression of SMOX by gender and stage (stage I compared to stage II, stage II compared to stage III). In the GSE42127 dataset, the only stage made a difference in SMOX expression (stage I compared to stage II). In the GSE68465 dataset, T, N, and stage all made a difference to SMOX expression (T1 compared to T2, N0 compared N1, stage I compared to stage II) (*Figure 2*) (11).

Functional enrichment analysis of differential genes in GSE68465

We obtained 389 differential genes from GSE68465, and the screening criteria were |log2FC| >0.5 and P value <0.05 (*Figure 3A,B*). We conducted GO enrichment analysis to understand further the location, function, and biological pathways of gene products in cells. The most enriched pathways were organelle fission, nuclear division, chromosome segregation, mitotic nuclear division, nuclear chromosome segregation, sister chromatid segregation, mitotic sister chromatid segregation, and DNA replication (*Figure 3C*). Also, we learned about related metabolic pathways through KEGG enrichment analysis, such as
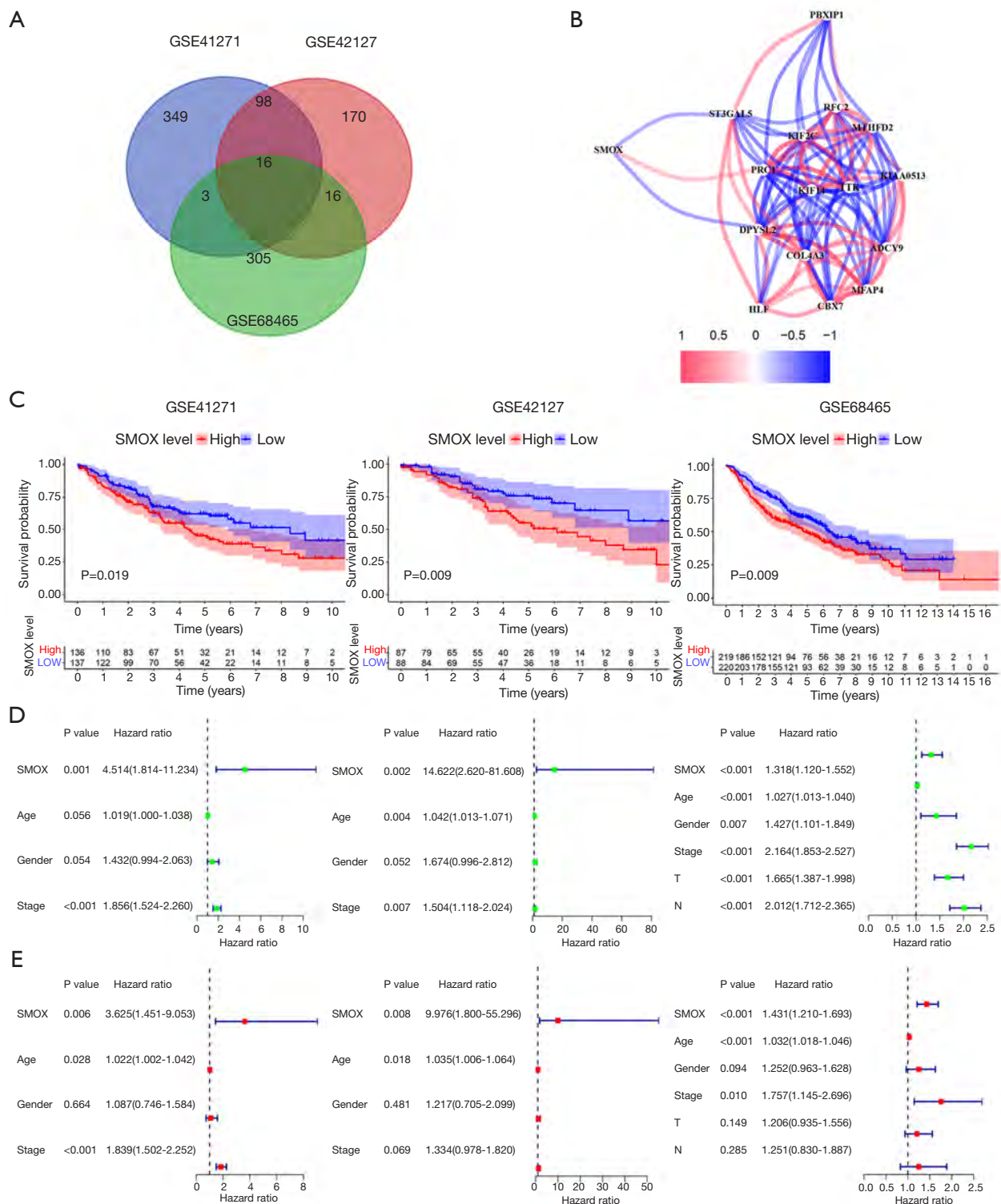
Figure 1 Selection and establishment of independent prognostic factors. (A) The Venn diagram was designed to reflect the common independent prognostic factors of the three GEO datasets. (B) The correlation network graph was used to show the association between 16 independent prognostic factors. The correlation threshold was 0.3. (C) Kaplan-Meier curves showed the survival difference between the high and low expression of SMOX in the three data sets (GSE41271, GSE42127, and GSE68465). (D,E) The forest plots showed the predictive power of SMOX for NSCLC in univariate and multivariate Cox regression analysis. GEO, Gene Expression Omnibus.

Page 6 of 17

Huang et al. SMOX as a prognostic biomarker of NSCLC

Table 1 Multivariate Cox proportional hazards model for SMOX in the study cohorts

SMOX	HR	HR.95L	HR.95H	P value
GSE41271	3.624725464	1.451240789	9.053380245	0.005826558
GSE42127	9.97648215	1.799957652	55.29585429	0.008471553
GSE68465	1.43135339	1.210274507	1.692816393	2.80E-05
TCGA	1.185147524	1.04800214	1.340240253	0.006785696
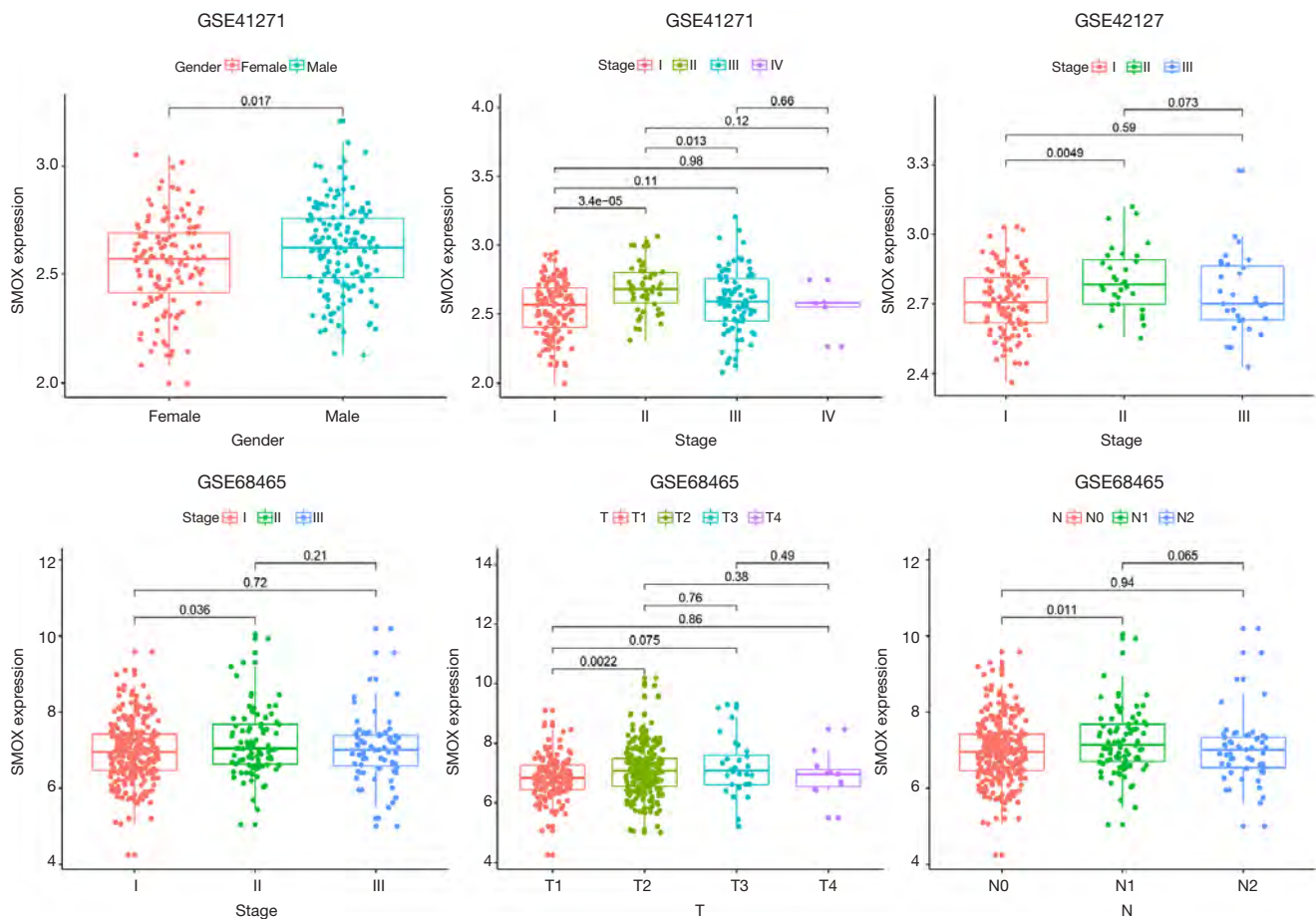
Figure 2 Clinical correlation analysis. The box plot showed the relationship between the clinical subgroups in the three GEO data sets (GSE41271, GSE42127, and GSE68465), including gender, stage staging and TNM staging. GEO, Gene Expression Omnibus.

cell cycle, phagosome, oocyte meiosis, DNA replication, and homologous recombination (*Figure 3D*). The most significant 60 differential genes were used to construct the protein interaction network and were visualized by Cytoscape7.2 software, including 27 up-regulated genes and 33 down-regulated genes (*Figure 3E*) (12).

Validation of SMOX in the TCGA Dataset

A total of 102 pairs of lung cancer and adjacent tissues in the TCGA database had significant differences (*Figure 4A*). Simultaneously, in the TCGA database, the expression of SMOX in 103 normal tissues and 999 lung cancer tissues
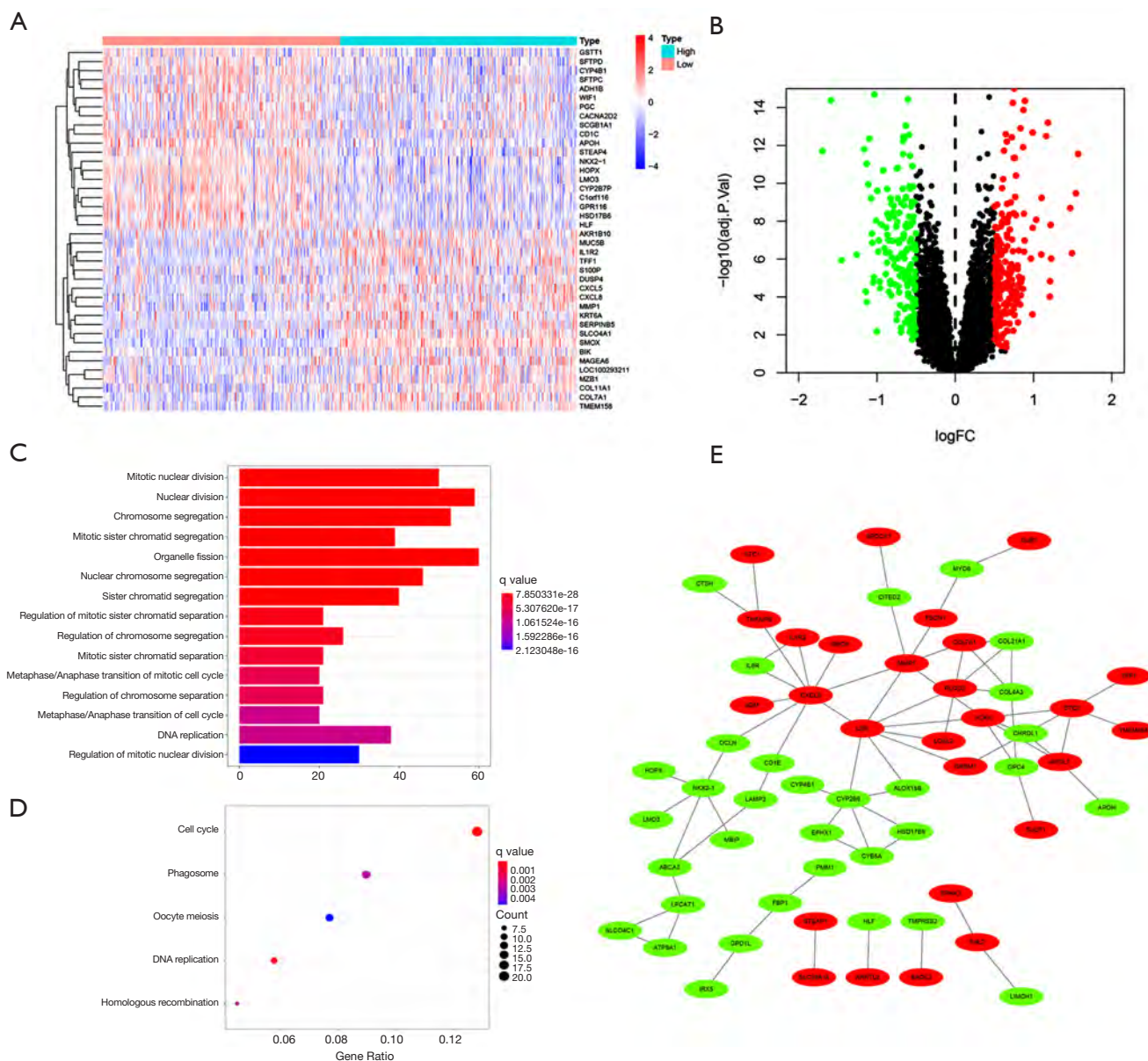
Figure 3 Difference analysis. (A) Using SMOX expression as the median value, patients were divided into high and low expression groups. The heat map was designed to reflect the expression of 40 distinct genes. Red represented high expression and blue represented low expression. (B) The volcano map was designed to depict the distribution of all meaningful differential genes. Red represented up-regulated genes, and green represented down-regulated genes. (C) The GO enrichment analysis aimed to reveal the annotation modules of differential genes. (D) The KEGG enrichment analysis aimed to display the enrichment pathways of differential genes. (E) Cytoscape 7.2 software visualized the 100 genes with the most obvious differences, and we obtained a protein interaction network constructed by 60 differential genes. Red represented up-regulated genes, and green represented down-regulated genes.

was significantly different (*Figure 4B*). The overall survival rate of SMOX in the low expression group was better than that of the high expression group (P=0.004) (*Figure 4C*). After univariate and multivariate Cox regression analyses,

we determined that SMOX was indeed an independent prognostic factor for lung cancer in the TCGA dataset (*Figure 4D*). Also, it was worth noting that the expression of SMOX differed significantly in terms of gender and T

Page 8 of 17

Huang et al. SMOX as a prognostic biomarker of NSCLC

stage (*Figure 4E*). GSEA contributed to showing the GO and KEGG pathways related to SMOX expression. In GO term annotation, 5 categories were positively related to the low expression group: regulation of presynapse organization, glucose catabolic process, wide pore channel activity, negative regulation of epidermis development, and glycolytic process through fructose 6 phosphate. Furthermore, 5 categories were negatively associated with the low expression group: branched-chain amino acid metabolic process, fatty acid beta-oxidation using acyl CoA oxidase, fatty acid catabolic process, immunoglobulin binding, and response to the tumor cell. Additionally, in the KEGG pathways, the 5 pathways that were positively correlated with the low expression group were small cell lung cancer, WNT signaling pathway, P53 signaling pathway, ERBB signaling pathway, and pathways in cancer. In comparison, the 5 pathways that were negatively related to the low expression group were FC epsilon RI signaling pathway, autoimmune thyroid disease, PPAR signaling pathway, GNRH signaling pathway, and vascular smooth muscle contraction (*Figure 4F*) (13-15).
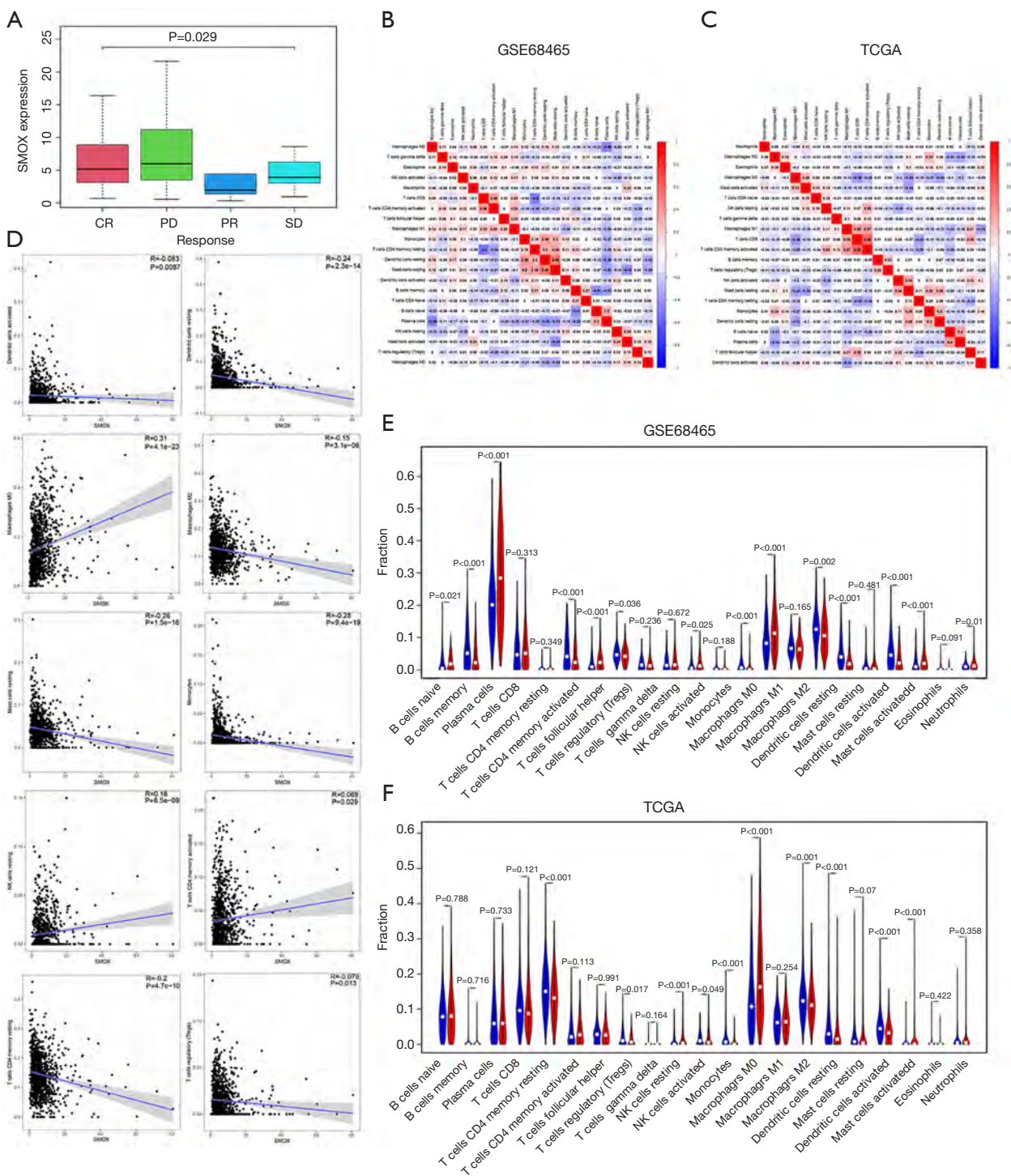
Immune infiltration analysis

We analyzed the prognostic efficacy of radiotherapy and chemotherapy patients in the TCGA database and found that there were significant differences between the drug efficacy indicators (*Figure 5A*). Since GSEA showed that SMOX was involved in some immune-related pathways, to explore the role of SMOX in the immune microenvironment, we conducted immune infiltration analysis in GSE68465, TCGA, respectively. In GSE68465, resting mast cells and resting dendritic cells showed the highest positive correlation, while CD4 memory resting T cells and CD8 T cells displayed the highest negative correlation (*Figure 5B*). In the TCGA data set, CD4 memory activated T cells and CD8 T cells displayed the highest positive correlation, while CD8 T cells and M0 macrophages showed the highest negative correlation (F*igure 5C*). Various immune cells showed a certain correlation with SMOX in the immune microenvironment, such as regulatory T cells, resting CD4 memory T cells, activated CD4 memory T cells, resting NK cells, monocytes, M0 Macrophages, M2 macrophages, resting mast cells, resting dendritic cells, and activated dendritic cells (*Figure 5D*). Furthermore, the distribution of the two groups in term of some immune cells was different. In GSE68465, naïve B cells, memory B cells, plasma cells,

CD4 memory resting T cells, CD4 memory activated T cells, follicular helper T cells, resting NK cells, monocytes, M0 macrophages, M2 macrophages, resting dendritic cells, resting mast cells, activated mast cells, and neutrophils showed significant differences between two groups (*Figure 5E*). CD4 memory resting T cells, regulatory T cells, resting NK cells, activated NK cells, Monocytes, M0 macrophages, M2 macrophages, resting dendritic cells, resting mast cells, and activated mast cells displayed differences between the two groups in the TCGA cohort (*Figure 5F*). It was worth noting that most of the immune cells with differences in GSE68465 and TCGA were the same (16,17). To further analyze the immune microenvironment of lung cancer, we calculated the immune and matrix scores of the patients and divided the patients into two groups according to the median value of the expression of SMOX. Significant differences were found in the immune and stromal scores between the two groups (*Figure 6A,B*). According to the median values of the immune score and matrix score, patients were divided into high and low-risk groups, and the top 50 differential genes with the most significant differences were shown in a heat map (*Figure 6C,D*). Secondly, the differential genes obtained through immune scores and matrix scores were crossed, and it was found that there were 1,222 differential up-regulated genes and 442 differential down-regulated genes (*Figure 6E*). The enrichment analysis of differential genes found that GO annotation was mainly enriched in immune response-activating cell surface receptor signal transduction and lymphocyte-mediated immunity, among others (*Figure 6F*). KEGG pathway enrichment analysis showed that the main enrichment was in cytokine-cytokine receptor interaction, staphylococcus aureus infection, and hematopoietic cell lineage, among others (*Figure 6G*). Costimulatory molecules are components involved in transmitting costimulatory signals, affecting the activation of T lymphocytes and B lymphocytes, which was fully verified in a violin plot. The expression level of costimulatory molecules was extracted from the GEO dataset and TCGA dataset, respectively. By dividing the costimulatory molecules into two groups according to the level of SMOX expression and comparing the differences between the two groups, in these 4 datasets, 9 costimulatory molecules were different between the two groups. They were CD40LG, HLA-DMA, HLA-DMB, HLA-DPA1, HLA-DPB1, HLA-DQB1, HLA-DQB2, HLA-DRA, and HLA-DRB6 (*Figure 7A,B,C,D*). Immune checkpoints refer to a series of molecules expressed on immune cells
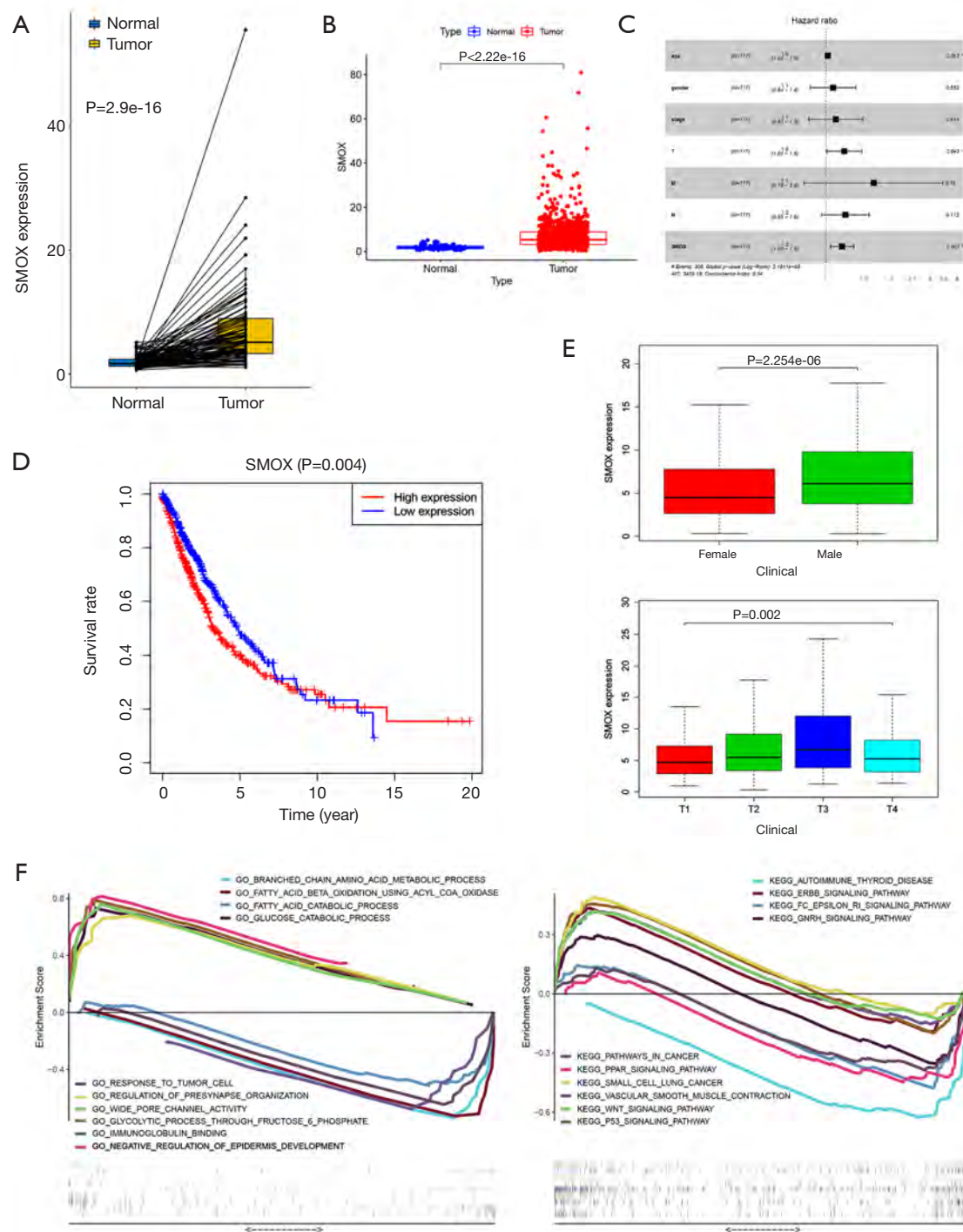
Figure 4 Validation of SMOX as an independent prognostic factor in the TCGA data set. (A) Paired difference analysis showed the difference of SMOX between cancer and adjacent cancer in the TCGA database. (B) The box plot showed the difference analysis of SMOX in normal and cancer tissues in the TCGA database. (C) The forest plot showed that SMOX was indeed an independent prognostic factor compared with other clinical factors in the TCGA database. (D) Kaplan-Meier curve showed the difference in survival between the high and low expression groups of SMOX. (E) Clinical correlation analysis showed the difference of SMOX in gender and T stage in the TCGA database. (F) GSEA enrichment analysis showed ten pathways with obvious enrichment in GO analysis and KEGG analysis, respectively. *P<0.05; **P<0.01. GO, Gene Ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes; TCGA, The Cancer Genome Atlas.

Page 10 of 17

Huang et al. SMOX as a prognostic biomarker of NSCLC

Figure 5 Immune infiltration analysis. (A) The expression of SMOX differed significantly among the drug efficacy. (B) The scatter plot showed the correlation between SMOX and 10 immune cells. (C,D) The correlation heat map displayed the correlation between SMOX and 22 immune cells in the GSE68465 and TCGA databases, respectively. (E,F) The violin chart indicated the difference among the immune cells in the high and low SMOX expression groups. CR, complete response; PR, partial response; PD, progressive disease; SD, stable disease; TCGA, The Cancer Genome Atlas.
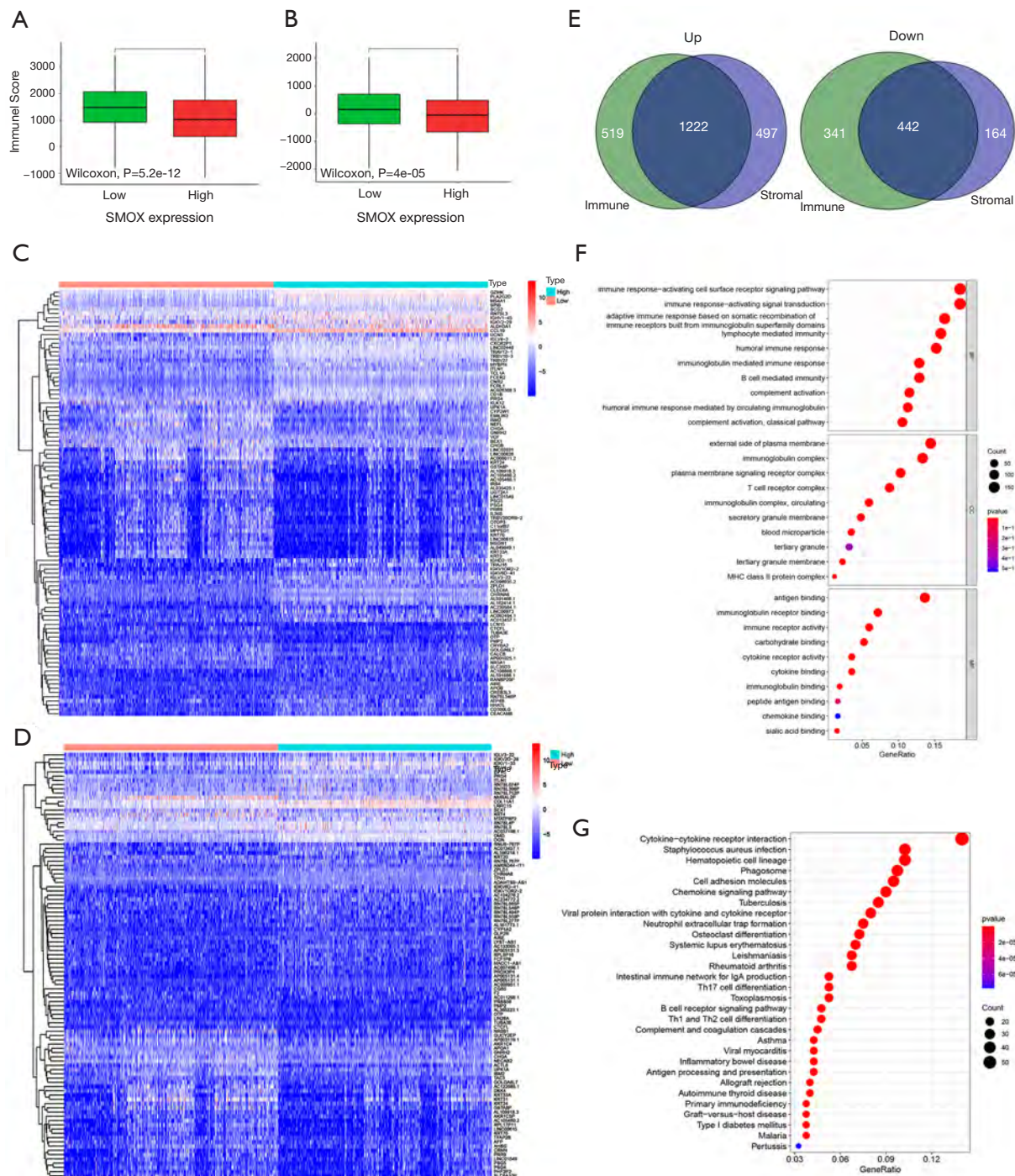
Figure 6 Immune microenvironment analysis. (A) The box plot revealed the difference in the immune score between the high and low expression groups of SMOX. (B) The box plot revealed the difference in matrix scores between the high and low expression groups of SMOX. (C) The heat map displayed the top 50 differential genes in the immune score between the high and low risk groups of SMOX expression. (D) The heat map showed the top 50 differential genes between the matrix scores in the high and low risk groups of SMOX expression. Red represented high expression and blue represented low expression. (E) The Venn diagram respectively showed the intersection of the up- and down-regulated differential genes between the immune score and the matrix score. (F,G) GO and KEGG enrichment analysis revealed the enrichment pathways and annotation modules of differential genes. GO, Gene Ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes.

Page 12 of 17

Huang et al. SMOX as a prognostic biomarker of NSCLC

regulating the degree of immune activation. They play an important role in preventing autoimmunity (abnormal immune function, attacking normal cells). We analyzed the differences in genes related to immune checkpoints in the SMOX high and low expression groups and reflected them in heat maps (*Figure 7E*). It is worth noting that there were significant differences between the high and low groups of SMOX expression in terms of TMB (*Figure 7F*). Eventually, the scores of immune cells and immune-related functions in each sample were obtained through SSGSEA analysis (18). In the correlation analysis of the immune score, SMOX expression in the high and low groups showed significant differences in terms of the correlation of immune cells (*Figure 7G*) and immune functions (*Figure 7H*).

External experimental verification

Predictions from the GEO and TCGA databases found that patients with low SMOX expression survive longer. It was speculated that the expression of SMOX in adjacent tissues was lower than that in tumor tissues. To further verify the expression of SMOX in NSCLC, western blot analysis was performed. It was found that SMOX had high-level expression in most of the 24 pairs of NSCLC tumor and paracancerous tissues, while it had low-level expression in paracancerous tissues (*Figure 8A,B*) (19). Notably, in 24 pairs of NSCLC tumor tissues and paracancerous tissues, SMOX had high-level expression in tumor tissues and low-level expression in normal tissues determined by RT-PCR analysis (*Figure 8C*) (20,21). Immunohistochemical analysis showed that SMOX was indeed highly expressed in 155 NSCLC patients (*Figure 8D*) (22). Furthermore, analysis of survival data from 155 NSCLC patients showed that there was a significant difference in the survival prognosis in the low-SMOX expression group, and patients with low SMOX expression had a better survival prognosis (*Figure 8E*). The median value of SMOX expression divided patients into high and low expression groups (Table S1). Finally, the forest plot showed that SMOX was identified as an independent prognostic factor for 155 NSCLC patients (*Figure 8F,G*).

In conclusion, SMOX can be used as a predictive target for NSCLC.

Discussion

In recent years, the incidence and mortality of lung cancer have increased rapidly. The 5-year survival rate of advanced lung cancer is less than 10%. Therefore, the early

detection of lung cancer is extremely important. Although there are different research directions, the ultimate goal is to diagnose and treat lung cancer. Basal cells are stem/progenitor cells that maintain airway homeostasis and repair after epithelial injury and are candidate source genes for lung squamous cell carcinoma (23), while alveolar type 2 (AT2) cells are the origin cells for lung adenocarcinoma (24). Circulating tumor cells (CTCs) are considered a valuable marker for early diagnosis, prognosis, and risk stratification in cancer patients, and they have been found to predict survival in patients with various types of cancer, including lung cancer (25). These studies aimed to understand the precursors of lung cancer and identify molecular changes during its development to facilitate early detection and better molecular targeting in the early stages of the disease.

The TME is mainly composed of many stromal cells, including immune cells, which are now considered key regulators of tumor development. According to the theory of cancer immunoediting, immune cells can kill tumor cells, especially in the early stages of tumor development, thereby reducing tumor incidence (26). Tumor evolution often leads to many heterogeneous cancer cell types in a single tumor, and heterogeneity includes epigenetic and gene expression changes, which cannot be identified by histological characteristics alone. Cancer cells invade nearby healthy tissues and the infiltration of immune cells, resulting in more complex structures of tumor and non-tumor cells (27). The extensive deletion of immunomodulatory genes is the reason for the low immune infiltration in cold immune tumors. The harmonious up-regulation of multiple immune checkpoint proteins may be due to the resistance of anti-programmed cell death protein 1 monotherapy on immune-hot tumors (28). Inflammatory cytokines in the TME promote immune cell infiltration. Infiltration of immune and peritumor stromal cells supports tumor growth, angiogenesis, metastasis, and immunosuppression through communication with inflammatory cytokines and cell adhesion molecules. Notably, the infiltrating immune cells and tumor cells have immunosuppressive molecules, such as programmed death ligand 1 (PD-L1) and CD80/CD86. Inhibition of cytotoxic T cells promotes tumor avoidance of immune surveillance and larger malignancies. Also, glycosylation and sialylation of proteins overexpressed on the surface of cancer cells have been shown to enhance immune escape and metastasis (29). The intricate interactions between malignant cells, host cells, and non-cellular components play key roles in different stages of tumor development, progression, and metastasis.
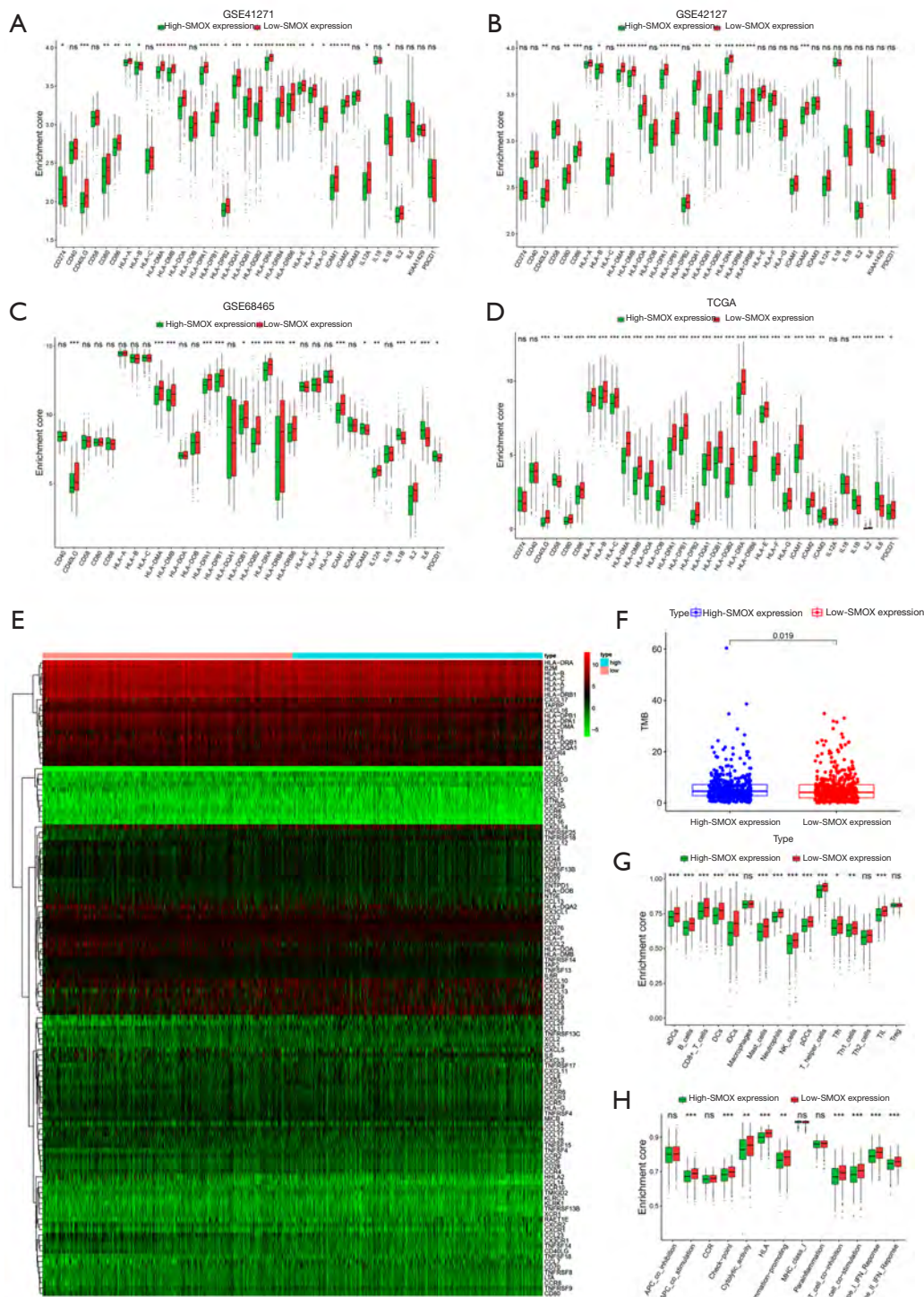
Figure 7 Immune correlation analysis. (A,B,C,D) Differences in the level of SMOX expression between costimulatory molecules in the four data sets (GSE41271, GSE42127, GSE68465, and TCGA). (E) The heat map indicated the difference between the high and low expression groups of SMOX between the immune check points. (F) The difference in TMB between high and low expression groups of SMOX. TMB is the total number of mutations per megabase in tumor tissue. (G,H) Differential expression of SMOX high and low expression group in SSGSEA analysis. *P<0.05; **P<0.01; ***P<0.001; ns, P≥0.05. TCGA, The Cancer Genome Atlas; TMB, tumor mutational burden.

Page 14 of 17

Huang et al. SMOX as a prognostic biomarker of NSCLC
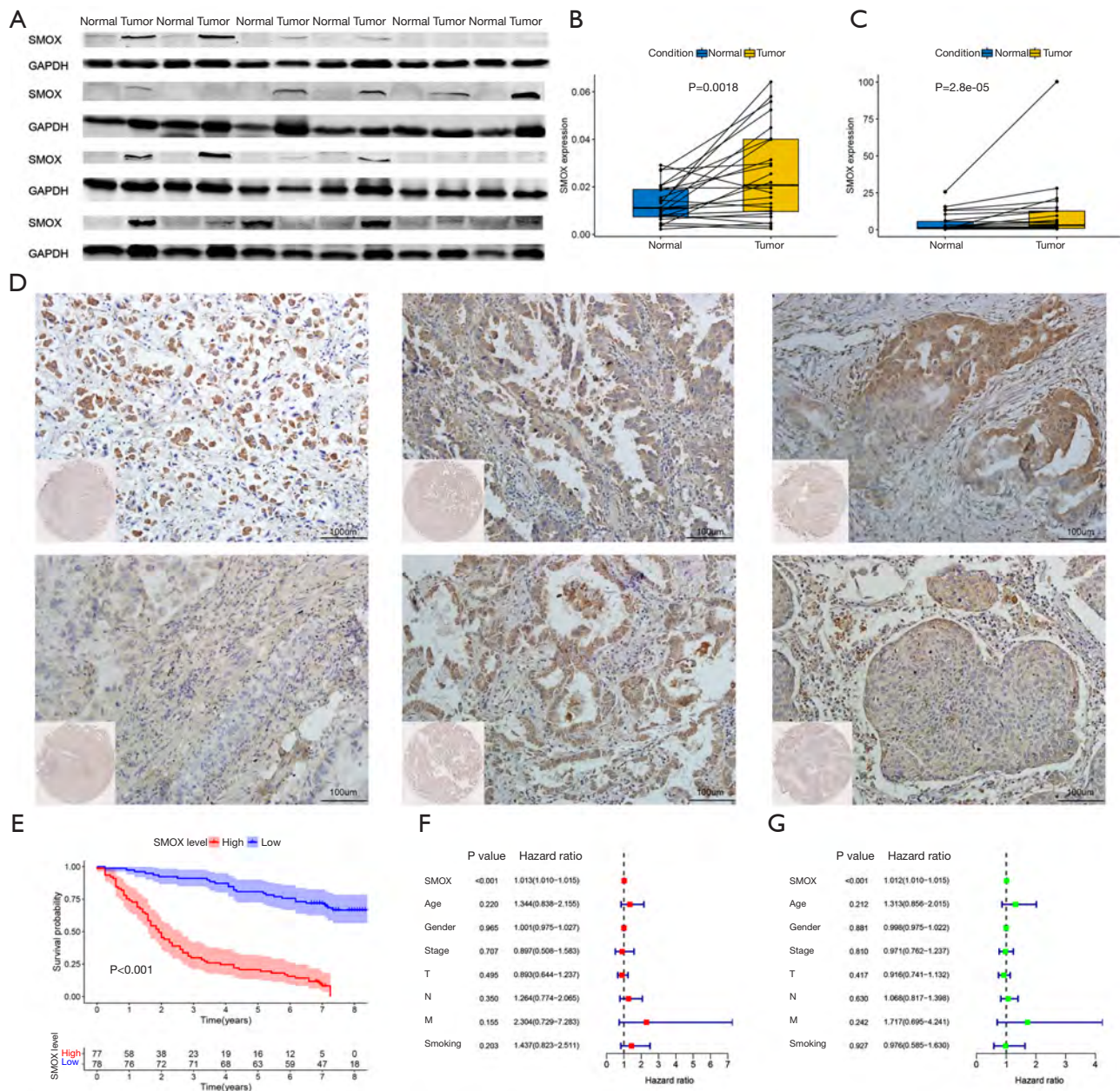
Figure 8 External experimental verification. (A,B) 24 Western blot analysis of tumors and adjacent tissues of NSCLC showed that SMOX was highly expressed in tumor tissues, but was low in adjacent tissues. (C) RT-PCR showed that SMOX was expressed in 13 pairs of NSCLC tumors and adjacent tissues. SMOX was highly expressed in tumor tissues and low in adjacent tissues. (D) Immunohistochemical analysis showed the expression of SMOX in tumor tissues. (E) Kaplan-Meier curves displayed the survival prognosis of 139 NSCLC patients in the high and low expression of SMOX. (F,G) The forest plots revealed the predictive power in univariate and multivariate Cox regression analysis. NSCLC, non-small cell lung cancer; RT-PCR, reverse transcription-polymerase chain reaction.

Tumor cells and stromal cells communicate through the secretion of signaling molecules such as integrins and other receptors and growth factors, cytokines, chemokines, and inflammatory mediators.

The signaling pathway mediated by chemokines has become the main mechanism for host cells to play multiple roles in tumor development. In response to tumor stimulation, chemokines derived from host cells further

activate the signaling cascade, supporting the ability of tumor cells to invade the surrounding basement membrane and extracellular matrix. Host-derived chemokines act on endothelial cells to increase their permeability and promote the infiltration and extravasation of tumor cells. The interaction between tumor cells and host neutrophils in the vascular system initiates chemokines to drive the recruitment of inflammatory cells, protecting CTCs from immune attack (30). Also, by stimulating the body's immune system, cancer immunotherapy has emerged as a promising strategy for treating many types of cancer. This therapy not only eradicates tumor cells by inducing a strong anti-tumor immune response but also prevents tumor recurrence (31). The metabolism of cancer cells is usually very different from that of normal cells. However, in tumor masses, the continuous interference and competition between different cells for nutrition and oxygen lead to metabolic changes in cancer cells and different stromal cells and immune cells in the TME. This is highly correlated with tumor progression (32). Human leukocyte antigen-DM (HLA-DM) is a component of the major histocompatibility complex class II (MHCII) antigen processing and presentation pathway.

HLA-DM shapes the immune system by differentially catalyzing peptide exchange on MHCII molecules to edit the peptide-MHCII (pMHCII) repertoire by imposing a bias on the foreign and self-derived peptide cargos which are presented on the cell surface for immune surveillance and tolerance induction by CD4+ T cells (33). The HLA complex contains the most polymorphic genes in the human genome. HLA class II genes define the specificity of the adaptive immune response. Hereditary variations in the *HLA* gene are associated with autoimmunity and susceptibility to infectious diseases (34). HLA-DPA1/DPB1 plays a key role in antigen presentation to CD4(+) helper T cells. Furthermore, inflammation can provide biologically active molecules for the TME, and the products of the inflammatory process can be considered as potential biomarkers. Studies have shown that neutrophil-to-lymphocyte ratio (NLR) and systemic immune-inflammation index (SII) are prognostic factor for NSCLC.

Corresponding studies have shown the ability of SMOX to produce oxidative cell stress, thereby driving cells into the initial stage of repair, which means SMOX induces cellular stress via an imbalance in ROS generation, but not apoptosis (35). Also, several studies have shown that SMOX plays an integral role in various diseases, but its role in NSCLC has not been reported. SMOX is associated with gastric inflammation and carcinogenesis and can be up-regulated in patients with inflammatory bowel disease. SMOX has an immunomodulatory role in experimental colitis polyamine flux (36). Simultaneously, SMOX promotes Helicobacter pylori induced cancer by causing inflammation, DNA damage, and activation of beta-catenin signaling (37). In naive mice, SMOX knockdown leads to airway hyperresponsiveness, remodeling, and apoptosis of bronchial epithelial cells (38). Also, SMOX is involved in diabetic retinal neurodegeneration and visual dysfunction (39). Related studies have shown that molecules with a high SMOX correlation have an important association with NSCLC. Polycomb group genes (PCGs) are epigenetic effectors, and PCGs are overexpressed in many cancers. They are a hot topic in cancer research and are essential for stem cell self-renewal and multi-performance (40). Dihydropyrimidinase-like proteins (DPYSLs) are a family of proteins that are developmentally regulated during nervous system maturation. Members of the DPYSL family are involved in cancer. Low expression of DPYSL1 in NSCLC is associated with adverse clinical outcomes and has a metastatic inhibitory effect (41).

Footnote

Reporting Checklist: The authors have completed the REMARK reporting checklist. Available at http://dx.doi.org/10.21037/atm-21-998

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at http://dx.doi.org/10.21037/atm-21-998). 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 appropriate investigated and resolved. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013). The study was approved by ethics board of Affiliated Hospital of Nantong University (No.: 2020-L002).

Page 16 of 17

Huang et al. SMOX as a prognostic biomarker of NSCLC

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. Dong RF, Zhu ML, Liu MM, et al. EGFR mutation mediates resistance to EGFR tyrosine kinase inhibitors in NSCLC: From molecular mechanisms to clinical research. Pharmacol Res 2021;167:105583.
2. van Veen S, Martin S, Van den Haute C, et al. ATP13A2 deficiency disrupts lysosomal polyamine export. Nature 2020;578:419-24.
3. Giraud J, Chalopin D, Blanc JF, et al. Hepatocellular Carcinoma Immune Landscape and the Potential of Immunotherapies. Front Immunol 2021;12:655697.
4. Zhang Y, Xu M, Sun Y, et al. Identification of LncRNAs Associated With FOLFOX Chemoresistance in mCRC and Construction of a Predictive Model. Front Cell Dev Biol 2021;8:609832.
5. Chen HQ, Chen DJ, Li Y, et al. DNA methylation and hydroxymethylation associated with gene expression regulatory network during 3-methylcholanthrene induced lung cell malignant transformation. Sci Total Environ 2021;771:144839.
6. Liu J, Hou S, Wang J, et al. A Prognostic 14-Gene Expression Signature for Lung Adenocarcinoma: A Study Based on TCGA Data Mining. Oxid Med Cell Longev 2020;2020:8847226.
7. Hu K, Li J, Wu G, et al. The novel roles of virus infection-associated gene CDKN1A in chemoresistance and immune infiltration of glioblastoma. Aging (Albany NY) 2021;13:6662-80.
8. Ni S, Li J, Qiu S, et al. KIF21B Expression in Osteosarcoma and Its Regulatory Effect on Osteosarcoma Cell Proliferation and Apoptosis Through the PI3K/AKT Pathway. Front Oncol 2021;10:606765.
9. Liang H, Xu M, Xiong Z, et al. Identification of miRNAs as diagnostic and prognostic markers in hepatocellular carcinoma. Aging (Albany NY) 2021;13:6115-33.
10. Song Y, Wang J, Sun J, et al. Screening of Potential Biomarkers for Gastric Cancer with Diagnostic Value Using Label-free Global Proteome Analysis. Genomics Proteomics Bioinformatics 2021. doi: 10.1016/j.gpb.2020.06.012.
11. Cai X, Qiu W, Qian M, et al. A Candidate Prognostic Biomarker Complement Factor I Promotes Malignant Progression in Glioma. Front Cell Dev Biol 2021;8:615970.
12. Yu K, Li D, Xu F, et al. IDO1 as a new immune biomarker for diabetic nephropathy and its correlation with immune cell infiltration. Int Immunopharmacol 2021;94:107446.
13. Yu F, Liang M, Wu W, et al. Upregulation of Long Non-Coding RNA GCC2-AS1 Facilitates Malignant Phenotypes and Correlated With Unfavorable Prognosis for Lung Adenocarcinoma. Front Oncol 2021;10:628608.
14. Wu WL, Li S, Zhao GJ, et al. Identification of circular RNAs as novel biomarkers and potentially functional competing endogenous RNA network for myelodysplastic syndrome patients. Cancer Sci 2021;112:1888-98.
15. Zhu P, Feng R, Lu X, et al. Diagnostic and prognostic values of AKR1C3 and AKR1D1 in hepatocellular carcinoma. Aging (Albany NY) 2021;13:4138-56.
16. Kim A, Wu X, Allende DS, et al. Gene deconvolution reveals aberrant liver regeneration and immune cell infiltration in alcohol-associated hepatitis. Hepatology 2021. doi: 10.1002/hep.31759.
17. Zeng H, Liu X, Zhang Y. Identification of Potential Biomarkers and Immune Infiltration Characteristics in Idiopathic Pulmonary Arterial Hypertension Using Bioinformatics Analysis. Front Cardiovasc Med 2021;8:624714.
18. Yan H, Chen Y, Yang Z, et al. An Immune Cell Signature Is Associated With Disease-Free Survival and Adjuvant Chemosensitivity of Patients With Resectable Gastric Cancer. Front Immunol 2021;11:621623.
19. Wang X, Xia X, Xu E, et al. Estrogen Receptor Beta Prevents Signet Ring Cell Gastric Carcinoma Progression in Young Patients by Inhibiting Pseudopodia Formation via the mTOR-Arpc1b/EVL Signaling Pathway. Front Cell Dev Biol 2021;8:592919.
20. Zhang C, Wang X, Fang D, et al. STK39 is a novel kinase contributing to the progression of hepatocellular carcinoma by the PLK1/ERK signaling pathway. Theranostics 2021;11:2108-22.
21. Vandierendonck A, Degroote H, Vanderborght B, et al. NOX1 inhibition attenuates the development of a pro-tumorigenic environment in experimental hepatocellular carcinoma. J Exp Clin Cancer Res 2021;40:40.
22. Fassan M, Milione M, Maddalena G, et al. Synaptophysin

expression in (V600EBRAF-)mutated advanced colorectal cancers identifies a new subgroup of tumours with worse prognosis. Eur J Cancer 2021;146:145-54.

23. Hynds RE, Janes SM. Airway Basal Cell Heterogeneity and Lung Squamous Cell Carcinoma. Cancer Prev Res (Phila) 2017;10:491-3.

24. Sainz de Aja J, Dost AFM, Kim CF. Alveolar progenitor cells and the origin of lung cancer. J Intern Med 2021;289:629-35.

25. Yousefi M, Ghaffari P, Nosrati R, et al. Prognostic and therapeutic significance of circulating tumor cells in patients with lung cancer. Cell Oncol (Dordr) 2020;43:31-49.

26. Strack E, Rolfe PA, Fink AF, et al. Identification of tumor-associated macrophage subsets that are associated with breast cancer prognosis. Clin Transl Med 2020;10:e239.

27. Pierce SE, Kim SH, Greenleaf WJ. Finding needles in a haystack: dissecting tumor heterogeneity with single-cell transcriptomic and chromatin accessibility profiling. Curr Opin Genet Dev 2021;66:36-40.

28. Huang C, Chen L, Savage SR, et al. Proteogenomic insights into the biology and treatment of HPV-negative head and neck squamous cell carcinoma. Cancer Cell 2021;39:361-79.e16.

29. Huang HW, Chang CC, Wang CS, et al. Association between Inflammation and Function of Cell Adhesion Molecules Influence on Gastrointestinal Cancer Development. Cells 2021;10:67.

30. Hussain S, Peng B, Cherian M, et al. The Roles of Stroma-Derived Chemokine in Different Stages of Cancer Metastases. Front Immunol 2020;11:598532.

31. Thakur N, Thakur S, Chatterjee S, et al. Nanoparticles as Smart Carriers for Enhanced Cancer Immunotherapy. Front Chem 2020;8:597806.

32. Virga F, Quirico L, Cucinelli S, et al. MicroRNA-Mediated Metabolic Shaping of the Tumor Microenvironment. Cancers (Basel) 2021;13:127.

33. Reyes-Vargas E, Barker AP, Zhou Z, et al. HLA-DM catalytically enhances peptide dissociation by sensing peptide-MHC class II interactions throughout the peptide-binding cleft. J Biol Chem 2020;295:2959-73.

34. Wittig M, Anmarkrud JA, Kassens JC, et al. Development of a high-resolution NGS-based HLA-typing and analysis pipeline. Nucleic Acids Res 2015;43:e70.

35. Fratini E, Cervelli M, Mariottini P, et al. Link between spermine oxidase and apoptosis antagonizing transcription factor: A new pathway in neuroblastoma. Int J Oncol 2019;55:1149-56.

36. Gobert AP, Al-Greene NT, Singh K, et al. Distinct Immunomodulatory Effects of Spermine Oxidase in Colitis Induced by Epithelial Injury or Infection. Front Immunol 2018;9:1242.

37. Sierra JC, Piazuelo MB, Luis PB, et al. Spermine oxidase mediates Helicobacter pylori-induced gastric inflammation, DNA damage, and carcinogenic signaling. Oncogene 2020;39:4465-74.

38. Jain V, Raina S, Gheware AP, et al. Reduction in polyamine catabolism leads to spermine-mediated airway epithelial injury and induces asthma features. Allergy 2018;73:2033-45.

39. Liu F, Saul AB, Pichavaram P, et al. Pharmacological Inhibition of Spermine Oxidase Reduces Neurodegeneration and Improves Retinal Function in Diabetic Mice. J Clin Med 2020;9:340.

40. Crea F, Paolicchi E, Marquez VE, et al. Polycomb genes and cancer: time for clinical application? Crit Rev Oncol Hematol 2012;83:184-93.

41. Tan F, Wahdan-Alaswad R, Yan S, et al. Dihydropyrimidinase-like protein 3 expression is negatively regulated by MYCN and associated with clinical outcome in neuroblastoma. Cancer Sci 2013;104:1586-92.

Table S1 SMOX expression grouping. In immunohistochemistry, the groups were grouped by the median value of SMOX expression

patient/id	futime/year	fustat	SMOX	gender	age	stage	T	N	M	smoking
1A3	2.25	death	123.18	female	58	Stage II	T2	N0	M0	non-smoker
1A5	8.166666667	alive	8.21	male	68	Stage III	T3	N1	M0	non-smoker
1A6	7.916666667	alive	11.72	male	60	Stage III	T3	N1	M0	non-smoker
1A7	7.916666667	alive	25.62	male	64	Stage III	T4	N1	M0	non-smoker
1B1	6.5	death	49.01	male	63	Stage II	T2	N1	M0	non-smoker
1B10	6.833333333	alive	41.60	female	63	Stage II	T1	N1	M0	non-smoker
1B2	7.166666667	death	45.20	male	62	Stage III	T2	N2	M0	smoker
1B4	4.25	death	243.51	male	51	Stage III	T2	N2	M0	smoker
1B5	5.916666667	death	149.13	male	64	Stage II	T2	N1	M0	smoker
1B7	0.916666667	death	144.36	male	62	Stage I	T1	N0	M0	smoker
1C10	6.916666667	alive	97.55	female	65	Stage III	T3	N2	M0	non-smoker
1C3	6.166666667	death	99.63	male	67	Stage III	T3	N1	M0	non-smoker
1C5	1.666666667	death	196.08	male	71	Stage II	T2	N1	M0	non-smoker
1C6	8.166666667	alive	3.41	male	64	Stage I	T1	N0	M0	smoker
1C9	6.833333333	alive	41.72	male	68	Stage I	T1	N0	M0	smoker
1D1	0.5	death	262.40	female	68	Stage I	T1	N0	M0	non-smoker
1D10	2.5	death	75.74	female	56	Stage IV	T2	N1	M1	non-smoker
1D2	0.833333333	death	91.37	male	67	Stage I	T2	N0	M0	non-smoker
1D4	1.833333333	death	218.98	female	42	Stage I	T2	N0	M0	non-smoker
1D5	7.083333333	alive	161.2	male	51	Stage II	T2	N0	M0	non-smoker
1D6	7.083333333	alive	65.89	female	59	Stage II	T4	N1	M0	non-smoker
1D8	7.083333333	alive	108.92	female	68	Stage I	T1	N0	M0	non-smoker
1D9	1.75	death	73.40	female	65	Stage II	T2	N0	M0	non-smoker
1E1	7.333333333	alive	74.91	male	60	Stage I	T1	N0	M0	non-smoker
1E2	4.25	death	165.90	male	51	Stage IV	T4	N1	M1	smoker
1E3	2	death	90.84	male	45	Stage II	T1	N1	M0	non-smoker
1E6	4.25	death	61.81	male	62	Stage II	T2	N1	M0	non-smoker
1E7	1.666666667	death	242.02	male	65	Stage II	T2	N1	M0	smoker
1E8	2.916666667	death	257.54	female	72	Stage I	T2	N0	M0	non-smoker
1F1	6.916666667	death	126.96	female	60	Stage II	T1	N1	M0	non-smoker
1F10	3.833333333	death	275.62	female	53	Stage II	T2	N1	M0	non-smoker
1F2	7.333333333	alive	25.56	male	63	Stage II	T1	N1	M0	non-smoker
1F3	8	alive	13.25	female	63	Stage II	T1	N1	M0	non-smoker
1F5	2.833333333	death	238.15	male	63	Stage III	T4	N0	M0	smoker
1F6	7.333333333	alive	22.69	male	55	Stage III	T1	N2	M0	non-smoker

Table S1 (*continued*)

patient/id	futime/year	fustat	SMOX	gender	age	stage	T	N	M	smoking
1F9	0.583333333	death	357.90	male	61	Stage II	T2	N1	M0	non-smoker
1G1	6.583333333	death	120.68	male	56	Stage III	T1	N2	M0	non-smoker
1G10	1.916666667	death	103.88	male	80	Stage I	T1	N0	M0	non-smoker
1G2	4.083333333	death	50.70	male	71	Stage III	T3	N1	M0	non-smoker
1G3	1.416666667	death	335.88	male	58	Stage II	T2	N1	M0	smoker
1G4	2.25	death	229.04	female	64	Stage IV	T4	N0	M1	non-smoker
1G5	1.25	death	291.48	male	65	Stage III	T4	N2	M0	non-smoker
1G6	4.333333333	death	48.61	female	56	Stage II	T1	N1	M0	non-smoker
1G8	7.166666667	alive	23.02	male	52	Stage II	T2	N1	M0	non-smoker
1G9	1.5	death	42.98	male	42	Stage II	T2	N1	M0	non-smoker
3A1	0.833333333	death	353.20	male	66	Stage I	T1	N0	M0	non-smoker
3A10	0.25	death	350	male	61	Stage II	T2	N1	M0	non-smoker
3A2	7.416666667	alive	16.51	female	62	Stage III	T2	N2	M0	non-smoker
3A4	8.333333333	alive	5.46	male	67	Stage I	T1	N0	M0	smoker
3A5	8.333333333	alive	5.26	female	75	Stage I	T2	N0	M0	non-smoker
3A6	0.583333333	death	219.72	male	68	Stage III	T1	N2	M0	non-smoker
3A7	0.916666667	death	23.42	male	76	Stage II	T2	N1	M0	non-smoker
3A9	3.5	death	14.97	male	73	Stage I	T1	N0	M0	non-smoker
3B1	7.833333333	alive	22.01	male	69	Stage III	T4	N0	M0	non-smoker
3B10	7.333333333	death	36.13	female	55	Stage III	T2	N2	M0	non-smoker
3B2	7.833333333	alive	9.67	female	67	Stage III	T2	N2	M0	non-smoker
3B3	1.916666667	death	44.46	male	76	Stage I	T1	N0	M0	non-smoker
3B4	8.333333333	alive	8.63	male	73	Stage III	T4	N2	M0	non-smoker
3B7	1	death	365.05	male	73	Stage II	T2	N1	M0	non-smoker
3B8	1.25	death	212.54	male	57	Stage III	T1	N2	M0	non-smoker
3B9	1.583333333	death	121.21	female	56	Stage I	T1	N0	M0	non-smoker
3C1	0.583333333	death	110.29	male	54	Stage III	T2	N2	M0	non-smoker
3C2	8.416666667	alive	3.69	female	64	Stage I	T2	N0	M0	non-smoker
3C4	0.75	death	144.93	male	56	Stage III	T4	N0	M0	non-smoker
3C5	1.416666667	death	105.45	male	36	Stage III	T1	N2	M0	non-smoker
3C7	8.083333333	alive	10.62	male	64	Stage II	T2	N0	M0	non-smoker
3D1	0.833333333	death	123.14	female	49	Stage III	T2	N2	M0	non-smoker
3D10	7.916666667	alive	5.06	female	74	Stage II	T2	N1	M0	non-smoker
3D2	7.083333333	death	62.64	male	71	Stage III	T2	N2	M0	smoker
3D3	1.5	death	238.53	male	78	Stage II	T3	N0	M0	non-smoker

Table S1 (*continued*)

patient/id	futime/year	fustat	SMOX	gender	age	stage	T	N	M	smoking
3D4	8.083333333	alive	6.82	female	56	Stage III	T2	N2	M0	non-smoker
3D5	8.083333333	alive	3.06	male	75	Stage I	T1	N0	M0	non-smoker
3D6	8.083333333	alive	5.09	male	62	Stage II	T2	N1	M0	non-smoker
3D7	1.083333333	death	199.49	male	80	Stage II	T1	N1	M0	non-smoker
3D8	7.833333333	alive	6.60	male	71	Stage II	T2	N1	M0	smoker
3E1	7.833333333	alive	76.79	male	60	Stage III	T4	N1	M0	non-smoker
3E10	2	death	165.18	female	57	Stage III	T2	N2	M0	non-smoker
3E2	2.333333333	death	313.57	male	68	Stage I	T1	N0	M0	non-smoker
3E3	1.25	death	244.59	male	77	Stage II	T1	N1	M0	non-smoker
3E4	6.583333333	death	116.70	male	64	Stage II	T2	N0	M0	smoker
3E6	3.25	death	112.88	male	61	Stage II	T2	N0	M0	non-smoker
3E7	2.583333333	death	144.78	female	66	Stage II	T2	N0	M0	non-smoker
3E8	1.166666667	death	76.00	male	71	Stage III	T4	N2	M0	smoker
3E9	7.916666667	alive	10.69	male	60	Stage I	T2	N0	M0	smoker
3F1	3.416666667	death	137.07	male	67	Stage II	T2	N1	M0	non-smoker
patient/id	futime/year	fustat	SMOX	gender	age	stage	T	N	M	smoking
3F10	3.666666667	death	58.37	female	70	Stage II	T2	N1	M0	non-smoker
3F2	5.833333333	death	88.43	female	55	Stage II	T1	N1	M0	non-smoker
3F3	7.25	death	79.05	male	76	Stage I	T1	N0	M0	non-smoker
3F4	8	alive	19.57	female	44	Stage III	T4	N0	M0	non-smoker
3F5	1.583333333	death	271.25	male	77	Stage II	T2	N1	M0	smoker
3F6	8	alive	10.13	male	62	Stage III	T3	N1	M0	non-smoker
3F7	2.666666667	death	170.84	male	55	Stage III	T3	N1	M0	non-smoker
3F8	0.583333333	death	356.66	male	66	Stage I	T2	N0	M0	smoker
3F9	1.583333333	death	229.09	male	70	Stage III	T4	N2	M0	non-smoker
3G10	5.75	death	67.88	female	61	Stage III	T4	N2	M0	non-smoker
3G2	7	death	149.99	female	67	Stage II	T1	N1	M0	non-smoker
3G3	8.166666667	alive	4.04	male	66	Stage III	T4	N1	M0	non-smoker
3G4	8.166666667	alive	37.61	female	62	Stage II	T1	N1	M0	non-smoker
3G5	8.166666667	alive	6.46	male	64	Stage III	T4	N1	M0	non-smoker
3G8	7.916666667	alive	39.62	female	58	Stage II	T2	N1	M0	non-smoker
3G9	8.25	alive	14.08	female	45	Stage III	T1	N2	M0	non-smoker
4A1	8.25	alive	4.98	female	61	Stage II	T2	N0	M0	non-smoker
4A10	0.25	death	48.63	male	53	Stage I	T2	N0	M0	non-smoker
4A2	5.5	death	52.68	male	80	Stage I	T1	N0	M1	non-smoker

Table S1 (*continued*)

patient/id	futime/year	fustat	SMOX	gender	age	stage	T	N	M	smoking
4A4	7.5	alive	43.19	male	55	Stage I	T1	N0	M0	smoker
4A5	0.25	death	149.03	male	52	Stage II	T1	N1	M0	smoker
4A7	7.5	alive	37.69	female	66	Stage I	T1	N0	M0	non-smoker
4A8	4.25	death	55.66	female	61	Stage III	T4	N1	M0	non-smoker
4A9	7.5	alive	75.89	male	55	Stage II	T2	N1	M0	non-smoker
4B1	5.25	death	40.92	female	65	Stage I	T2	N0	M0	non-smoker
4B2	2.583333333	death	145.97	male	58	Stage III	T1	N2	M0	non-smoker
4B3	7.5	alive	26.03	male	74	Stage II	T2	N0	M0	non-smoker
4B4	5.083333333	death	119.51	male	72	Stage II	T2	N0	M0	non-smoker
4B5	6.666666667	alive	51.83	male	57	Stage II	T2	N1	M0	non-smoker
4B6	2.833333333	death	114.09	male	73	Stage III	T2	N2	M0	smoker
4B7	6.75	alive	50.06	female	55	Stage I	T1	N0	M0	smoker
4C1	6.75	alive	28.24	male	51	Stage II	T2	N0	M0	smoker
4C10	0.916666667	death	145.82	male	43	Stage I	T1	N0	M0	non-smoker
4C2	6.666666667	alive	55.65	male	50	Stage II	T2	N0	M0	non-smoker
4C3	6.166666667	death	45.78	male	90	Stage II	T3	N0	M0	non-smoker
4C4	6.666666667	alive	68.08	male	56	Stage I	T1	N0	M0	non-smoker
4C5	0.25	death	306.11	female	62	Stage I	T1	N0	M0	non-smoker
4C6	2	death	152.03	male	44	Stage II	T3	N0	M0	non-smoker
4C7	0.666666667	death	142.84	male	67	Stage II	T2	N1	M0	smoker
4C8	1.916666667	death	132.26	male	77	Stage II	T3	N0	M0	non-smoker
4D1	1.333333333	death	142.98	male	73	Stage I	T2	N0	M0	non-smoker
4D10	6.833333333	alive	42.38	female	63	Stage I	T1	N0	M0	non-smoker
4D2	4.083333333	death	60.64	female	57	Stage II	T2	N1	M0	non-smoker
patient/id	futime/year	fustat	SMOX	gender	age	stage	T	N	M	smoking
4D3	7	alive	46.39	male	65	Stage I	T2	N0	M0	non-smoker
4D4	6.166666667	death	57.58	female	77	Stage IV	T4	N1	M1	non-smoker
4D5	0.25	death	145.18	female	62	Stage I	T1	N0	M0	non-smoker
4D6	7	alive	115.83	female	58	Stage I	T1	N0	M0	non-smoker
4D7	1.833333333	death	168.73	female	42	Stage I	T2	N0	M0	non-smoker
4D8	0.416666667	death	341.69	male	63	Stage I	T1	N0	M0	smoker
4D9	6.833333333	alive	122.57	female	56	Stage II	T2	N1	M0	non-smoker
4E1	0	death	84.49	female	43	Stage I	T1	N0	M0	non-smoker
4E10	2.333333333	death	118.05	male	52	Stage I	T1	N0	M0	non-smoker
4E4	5.416666667	death	119.91	female	67	Stage II	T2	N1	M0	non-smoker

Table S1 (*continued*)

Table S1 (*continued*)

patient/id	futime/year	fustat	SMOX	gender	age	stage	T	N	M	smoking
4E5	5.166666667	death	45.53	female	49	Stage I	T1	N0	M0	non-smoker
4E6	2.666666667	death	229.82	male	63	Stage I	T2	N0	M0	non-smoker
4E9	7.083333333	alive	79	female	69	Stage III	T4	N0	M0	non-smoker
4F1	7.083333333	alive	64.88	male	64	Stage II	T2	N0	M0	smoker
4F3	7.083333333	alive	59.44	male	57	Stage I	T1	N0	M0	smoker
4F4	7.083333333	alive	29.66	male	69	Stage II	T2	N2	M0	non-smoker
4F6	1.666666667	death	123.45	male	75	Stage I	T1	N0	M0	smoker
4F7	4.416666667	death	103.18	male	74	Stage III	T2	N2	M0	non-smoker
4F8	3.416666667	death	72.83	female	57	Stage I	T1	N1	M0	non-smoker
4G1	2.083333333	death	132.77	male	67	Stage I	T1	N0	M0	non-smoker
4G2	1.583333333	death	123.89	female	58	Stage I	T2	N0	M0	non-smoker
4G3	6.916666667	alive	82.32	male	72	Stage I	T2	N0	M0	non-smoker
4G4	0.583333333	death	140.00	male	57	Stage III	T2	N2	M0	non-smoker
4G5	6.916666667	alive	30.24	male	73	Stage II	T2	N1	M0	non-smoker
4G7	7.083333333	alive	46.48	male	66	Stage I	T1	N0	M0	smoker
4G8	3.333333333	death	124.32	male	57	Stage I	T2	N0	M0	non-smoker
4G9	7.416666667	alive	65.60	female	57	Stage I	T1	N0	M0	non-smoker