



Novel GIRlncRNA Signature for Predicting the Clinical Outcome and Therapeutic Response in NSCLC

Qiangzhe Zhang¹, Xicheng Liu², Zhinan Chen³ and Sihe Zhang^{4*}

¹State Key Laboratory of Medicinal Chemical Biology and College of Pharmacy, Tianjin Key Laboratory of Molecular Drug Research, Nankai University, Tianjin, China, ²Department of Physiology and Pathophysiology, School of Basic Medical Sciences, Capital Medical University, Beijing, China, ³National Translational Science Center for Molecular Medicine, Department of Cell Biology, State Key Laboratory of Cancer Biology, Fourth Military Medical University, Xi'an, China, ⁴Department of Cell Biology, School of Medicine, Nankai University, Tianjin, China

OPEN ACCESS

Edited by:

Hongtao Xiao,
University of Electronic Science and
Technology of China, China

Reviewed by:

Yueguo Li,
Tianjin Medical University Cancer
Institute and Hospital, China
Yaojiang Huang,
Minzu University of China, China
Nishant Karadkhelkar,
The Scripps Research Institute,
United States

*Correspondence:

Sihe Zhang
sihezhang@nankai.edu.cn
<https://orcid.org/0000-0002-8923-1993>

Specialty section:

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

Received: 06 May 2022

Accepted: 23 June 2022

Published: 03 August 2022

Citation:

Zhang Q, Liu X, Chen Z and Zhang S
(2022) Novel GIRlncRNA Signature for
Predicting the Clinical Outcome and
Therapeutic Response in NSCLC.
Front. Pharmacol. 13:937531.
doi: 10.3389/fphar.2022.937531

Background: Non-small cell lung cancer (NSCLC) is highly malignant with driver somatic mutations and genomic instability. Long non-coding RNAs (lncRNAs) play a vital role in regulating these two aspects. However, the identification of somatic mutation-derived, genomic instability-related lncRNAs (GIRlncRNAs) and their clinical significance in NSCLC remains largely unexplored.

Methods: Clinical information, gene mutation, and lncRNA expression data were extracted from TCGA database. GIRlncRNAs were screened by a mutator hypothesis-derived computational frame. Co-expression, GO, and KEGG enrichment analyses were performed to investigate the biological functions. Cox and LASSO regression analyses were performed to create a prognostic risk model based on the GIRlncRNA signature (GIRlncSig). The prediction efficiency of the model was evaluated by using correlation analyses with mutation, driver gene, immune microenvironment contexture, and therapeutic response. The prognostic performance of the model was evaluated by external datasets. A nomogram was established and validated in the testing set and TCGA dataset.

Results: A total of 1446 GIRlncRNAs were selected from the screen, and the established GIRlncSig was used to classify patients into high- and low-risk groups. Enrichment analyses showed that GIRlncRNAs were mainly associated with nucleic acid metabolism and DNA damage repair pathways. Cox analyses further identified 19 GIRlncRNAs to construct a GIRlncSig-based risk score model. According to Cox regression and stratification analyses, 14 risk lncRNAs (AC023824.3, AC013287.1, AP000829.1, LINC01611, AC097451.1, AC025419.1, AC079949.2, LINC01600, AC004862.1, AC021594.1, MYRF-AS1, LINC02434, LINC02412, and LINC00337) and five protective lncRNAs (LINC01067, AC012645.1, AL512604.3, AC008278.2, and AC089998.1) were considered powerful predictors. Analyses of the model showed that these GIRlncRNAs were correlated with somatic mutation pattern, immune microenvironment infiltration, immunotherapeutic response, drug sensitivity, and survival of NSCLC patients. The GIRlncSig risk score model demonstrated good predictive performance (AUCs of ROC for 10-year survival was 0.69) and prognostic value

in different NSCLC datasets. The nomogram comprising GIRlncSig and tumor stage exhibited improved robustness and feasibility for predicting NSCLC prognosis.

Conclusion: The newly identified GIRlncRNAs are powerful biomarkers for clinical outcome and prognosis of NSCLC. Our study highlights that the GIRlncSig-based score model may be a useful tool for risk stratification and management of NSCLC patients, which deserves further evaluation in future prospective studies.

Keywords: long non-coding RNA, somatic mutation, genomic instability, prognostic signature, non-small cell lung cancer

INTRODUCTION

Lung cancer is the leading cause of cancer death worldwide, with an approximate 1.8 million deaths each year (Sung et al., 2021). About 85% of lung cancer patients are diagnosed with non-small cell lung cancer (NSCLC), of which lung adenocarcinoma (LUAD) and lung squamous cell carcinoma (LUSC) are the two major subtypes (Herbst et al., 2018). Although clinical approaches have achieved significant advances in NSCLC treatment, the 5-year survival rate is only 25% in 2021 (Sung et al., 2021), and there is an urgent need for the identification of novel prognostic biomarkers for improved risk stratification and enhanced therapeutic efficiency of NSCLC (Gridelli et al., 2015; Chen and Dhahbi, 2021; Peng et al., 2021).

Genomic instability and somatic mutations are two hallmarks of cancer and contribute essentially to malignant transformation (Hanahan and Weinberg, 2011). As predicted by the “Mutator Phenotype” hypothesis, mutations of DNA repair genes, oncogenes, and tumor suppressor genes (TSG) can cause increased genomic instability, which drives cancer onset and progression (Negrini et al., 2010). Additionally, defects in genes controlling chromosome cohesion, mitotic kinetochore-microtubule attachment, centrosome copy number, checkpoint function, and cell-cycle regulation can accelerate the genomic instability. Beyond that, chromosomal instability such as translocations, deletions, insertions, amplifications, and inversions of large segments as well as gains or losses of whole chromosomes can also cause genomic instability (Thompson et al., 2010; Bastians, 2015). Furthermore, epigenetic modifications, like DNA modification, histone variants and modifications, nucleosome remodeling, and non-coding RNA, together play important roles in keeping genomic stability (Reis et al., 2016; Feng and Riddle, 2020). Moreover, an undesirable tumor microenvironment can also increase genomic instability (Rummel et al., 2012), of which hypoxia is a major factor (Sonugur and Akbulut, 2019). Notably, genomic instability contributes to the acquisition of multidrug resistance in malignancy (Lee et al., 2011; Osrodek and Wozniak, 2021), which frequently gives rise to poor therapeutic response and patient outcome (Lukow et al., 2021). Therefore, novel biomarkers correlated with genomic instability are critical for cancer diagnosis, treatment, and prognosis.

Emerging evidence indicates that long noncoding RNAs (lncRNAs) play vital roles in regulating genomic stability (Nair et al., 2020). Multiple studies have revealed that lncRNAs can

preserve genomic stability in the process of DNA damage response and repair. For instance, a colorectal cancer-overexpressed oncogenic lncRNA, CRNDE, could reduce DNA damage and cell apoptosis after oxaliplatin treatment (Gao et al., 2017). Furthermore, lncRNA LINP1 could serve as a scaffold linking Ku80 and DNA-PKcs and consequently coordinate non-homologous end-joining pathways to enhance the repair of DNA double-strand breaks (Zhang et al., 2016). A poorly characterized lncRNA, NORAD, maintains genomic stability by sequestering the PUMILIO protein. Once NORAD was missing, PUMILIO induce chromosomal instability by hyperactively inhibiting mitosis, DNA replication, and DNA damage repair (Lee et al., 2016). These studies strongly suggest that genomic instability-related lncRNAs (GIRlncRNAs) may provide a vital molecular signature for predicting the malignant phenotype. Recently, Bao et al identified a genomic instability-related lncRNA signature (GIRlncSig) for improved prediction of breast cancer outcome, which combined lncRNA expression and the somatic mutation profile (Bao et al., 2020). Geng and Peng have identified two sets of GIRlncSig in LUAD and early LUAD with the favorable prognostic outcome (Geng et al., 2021; Peng et al., 2021). Although these GIRlncSigs have been associated with the prognosis of particular subtypes of lung cancer, GIRlncSig-associated tools for NSCLC prognosis have yet to be established. More importantly, the clinical significance and biological function of GIRlncSig in NSCLC remain largely unexplored.

In this study, we identified 19 somatic mutation-derived GIRlncRNAs in NSCLC by using the mutator hypothesis-derived computational frame. A GIRlncSig-based risk score model with reliable prognostic performance was constructed, which can be a powerful indicator of genomic instability, immune microenvironment infiltration, therapeutic response, drug resistance, and patient stratification, thereby improving the personalized treatment of NSCLC.

MATERIALS AND METHODS

Data Collection and Preprocessing

The overall procedure in this study is outlined as the roadmap (Supplementary Figure S1). Transcriptomic and clinical information on NSCLC (LUAD and LUSC) were downloaded from The Cancer Genome Atlas (TCGA) database via the UCSC Xena Browser (<https://xenabrowser.net/>). Somatic mutation data

TABLE 1 | Clinical information of NSCLC patients.

Covariate	Type	TCGA set (n = 975)	Training set (n = 683)	Testing set (n = 292)	p-value ^a
Gender	Male	584 (59.9%)	417 (61.05%)	167 (57.19%)	0.2597
	Female	391 (40.1%)	266 (38.95%)	125 (42.81%)	
Age (years)	>60	698 (71.59%)	489 (71.60%)	209 (52.74%)	0.9574
	≤60	262 (26.87%)	184 (26.93%)	78 (45.55%)	
	NA	15 (1.54%)	10 (1.46%)	5 (1.71%)	
Tumor stage	Stage I-II	772 (79.18%)	541 (79.21%)	231 (79.11%)	0.8944
	Stage III-IV	192 (19.69%)	135 (19.76%)	57 (19.52%)	
	Unknown	11 (1.13%)	7 (1.02%)	4 (1.37%)	
Pathologic M	M0	723 (74.15%)	517 (75.7%)	206 (70.55%)	0.1359
	M1	31 (3.18)	23 (3.37%)	8 (2.74%)	
Pathologic N	NX	221 (22.67%)	143 (20.93%)	78 (26.71%)	0.2510
	N0	626 (64.21%)	441 (64.57%)	185 (63.36%)	
	N1-3	333 (34.15%)	228 (33.38%)	105 (35.96%)	
Pathologic T	TX	16 (1.64%)	14 (2.05%)	2 (0.68%)	0.4256
	T1-2	821 (84.21%)	571 (83.6)	250 (85.62%)	
	T3-4	151 (15.49%)	109 (15.96%)	42 (14.38%)	
	TX	3 (0.31%)	3 (0.44%)	0 (0%)	

^aChi-squared test, $p < 0.05$ means significantly different.

were downloaded from TCGA database (<https://portal.gdc.cancer.gov/>). Counts data were used for the transcriptome data, and the muTect version was used for the somatic mutation data. The human gtf file containing the gene symbol was downloaded from the Ensembl database (Homo_sapiens.GRCh38.99.gtf.gz; <http://www.ensembl.org>). Transcriptomic, somatic mutation data, and clinical information were matched according to the sample name, and samples with missing data were excluded. Finally, 975 complete samples including gene expression, mutation, patient's survival, and other clinical variables were obtained. The mRNAs and lncRNAs of 975 samples were annotated based on the gtf file containing the gene symbol. Then, these samples were randomly distributed into training and testing sets at a ratio of 7:3 using the "caret" package in R. The training set of 683 patients was used to identify the GIRlncSig and construct the prognostic risk model. The testing set of 292 patients was used to validate the performance of our risk model. The clinical information of NSCLC patients is summarized in **Table 1**.

Identification of GIRlncRNAs

The lncRNA expression profile and somatic mutation pattern of 975 patients were combined to identify GIRlncRNAs by a mutator hypothesis-derived computational frame (Bao et al., 2020) (**Supplementary Figure S2**). Briefly, NSCLC patients were ranked in increasing order according to the cumulative number of somatic mutations. The top 25% of patients with low mutation frequency were designated as the genomic stable (GS) group, and the last 25% of patients with high mutation frequency were designated as genomic unstable (GU) group. lncRNA expression profiles between the two groups were compared, and a volcano plot was made using the "edgeR" package in R. Differentially expressed (DE) lncRNAs ($|\text{Fold Change}| > 1.0$ and adjusted $p < 0.05$) were defined as GIRlncRNAs, and their expression levels were normalized to all patients.

Co-Expression Network, GO and KEGG Enrichments, and Alternative Splicing Analysis

The mRNA-interacting GIRlncRNAs were extracted from the RNAInter database (<http://www.rna-society.org/rnainter/>). Then, the co-expression network of mRNA-interacting GIRlncRNAs was visualized by Cytoscape (V3.7.2) (Chin et al., 2014). To explore the biological function of GIRlncRNAs, GO functional enrichment and KEGG pathway analyses were performed using the "clusterProfiler," "org.Hs.eg.db," "enrichplot," and "ggplot2" packages in R. Statistical significance was considered with adjusted p -value < 0.05 . To determine the AS events associated with GIRlncRNAs, AS data of NSCLC were downloaded from LncAS2Cancer database (<https://lncrna2as.cd120.com/>), and integrated with corresponding GIRlncRNAs. Statistical histogram and Upset plot were drawn using "ggplot2" and "UpSetR" packages in R, respectively.

Hierarchical Cluster Analysis

The normalized DE-lncRNAs from 975 samples were collected for hierarchical cluster analysis using the pam method, and the spearman distances were calculated using the "ConsensusClusterPlus" package in R. All samples were divided into two clusters based on the spearman distances. One cluster with low mutation counts was assigned as GS-like subtype. The other with high mutation counts were defined as GU-like subtype (Mann-Whitney U test, $p < 0.05$). Finally, Kaplan-Meier survival curves and the expression heatmap of DE-lncRNAs in the two subtypes were plotted using "survival," "survminer," and "ComplexHeatmap" packages in R.

GIRlncRNA-Clustered Molecular Subtype and Characteristic Analyses

To estimate the differential mutation frequency between GS and GU-like subtypes of NSCLC patients, driver genes from Cancer

Gene Census (CGC) catalog in the COSMIC database (<https://cancer.sanger.ac.uk/census>) were downloaded. Fifty-four driver genes with corresponding MAF files were extracted and determined their mutation frequency. The landscape of mutation frequency was drawn using the “maftools” package in R. Furthermore, the differential expression of driver genes from two NSCLC subtypes was analyzed using “edgeR” package in R. The cutoff criteria were $|\log_2 \text{FC}| > 1$ and adjusted $p < 0.05$. The Wilcoxon test was used to analyze the differentially expressed GIRlncRNAs. Moreover, the immune, stromal, and ESTIMATE scores for the two NSCLC subtypes were determined by the “estimate” package in R. In addition, tumor mutation burden (TMB) data were downloaded from TCGA database (<https://gdc.cancer.gov/>) and TMB scoring for the two NSCLC subtypes was performed. Finally, the scores for the mRNA expression-based stemness index (mRNAsi) were calculated, and NSCLC patients were stratified into different molecular subtypes by referring to the matrix in Malta et al. (2018). Boxplots for GS- and GU-like groups were drawn using the “ggpubr” package in R (Student t-test, adjusted $p < 0.05$).

Identification of GIRlncSig and Performance Evaluation

To estimate the correlation of GIRlncRNA expression with overall survival (OS) of NSCLC patients, a univariate Cox proportional hazards regression analysis was performed in the training set by using “survival” and “survminer” packages in R. The candidate GIRlncRNAs were screened with $p < 0.01$. Then, the least absolute shrinkage and selection operation (LASSO) Cox regression analysis was carried out to identify the GIRlncRNAs with the most robust prognostic values. Finally, the resulting GIRlncRNAs were collected to construct a GIRlncSig based on the weighted expression level and coefficient (coef) from LASSO regression analysis. The risk score formula for GIRlncRNAs was calculated as follows:

$$\text{GIRlncSig Risk Score}_i = \sum_{j=1}^n \exp_{ji} \times \beta_j.$$

The “ \exp ” and “ β ” represent the expression level and coefficient of each prognostic lncRNA, respectively. GIRlncSig risk score represents the sum of expression level of each prognostic lncRNA multiplied by the coef of corresponding GIRlncSig. i represents the sample, and j represents the prognostic lncRNA. Based on the risk score for all samples, high and low-risk patients were then recognized by using the median value as the cut-off point. Evaluation analyses, including the Kaplan–Meier survival curve, receiver operating characteristic curve (ROC), risk distribution, survival status of all patients, and heatmap of selected GIRlncSig expression profile were applied to test the predication performance of our risk model in training, testing, and TCGA sets.

Independent Prognostic and Clinical Stratification Analysis

To test whether the GIRlncSig risk score could be potentiated as a prognostic factor independently from other clinical variables

(age, gender, and tumor stage), univariate and multivariate Cox regression analyses (UCRA and MCRA) were performed. All variables with independent prognostic values were selected from TCGA database when their p -values were less than 0.05. To test the prognostic stability of GIRlncSig scoring, a clinical stratification analysis was conducted. Patients were first divided into subgroups according to clinical variables, including age (≤ 60 and > 60), gender (female and male), pathologic T (T1–T2 and T3–T4), and tumor stage (I–II and III–IV). Second, patients in each subgroup were ranked according to their GIRlncSig risk score, and assigned to high or low-risk subgroups by referring to the median value. Finally, the survival difference between high and low-risk subgroups was calculated (log-rank test, $p < 0.05$).

Correlation Analysis Between GIRlncSig and Drive Genes or the Microenvironment

Expression profiles of GIRlncSig and driver genes of NSCLC were extracted, and their correlation was calculated using “ggplot2” and “ggpubr” packages in R. The cutoff criteria were $R > 0.3$ and $p < 0.05$, respectively. Correlation between GIRlncSig score and immune, or stromal, or ESTIMATE, or mRNAsi or TMB scores were further calculated ($p < 0.05$).

Immune Cell Infiltration, Checkpoint Inhibitor-Related Genes, and Therapeutic Response

The mRNA expression matrix of each patient was converted to 22 types of immune cell matrix with cutoff criteria of $p < 0.05$. Tumor-infiltrating immune cells include naive B cells, memory B cells, plasma cells, CD8⁺ T cells, naive CD4⁺ T cells, resting memory CD4⁺ T cells, activated memory CD4⁺ T cells, follicular helper T cells, regulatory T cells (Tregs), gamma delta T cells, resting dendritic cells, activated dendritic cells, monocytes, macrophages M0, M1, M2, resting natural killer (NK) cells, activated NK cells, resting mast cells, activated mast cells, eosinophils, and neutrophils. Based on their association with clinical outcomes, NSCLC patients were classified into high- and low-risk groups using the Cell type Identification by Estimating Relative Subsets of RNA Transcript (CIBERSORT) algorithm (Chen et al., 2018). The type and distribution of infiltrating immune cells were analyzed by “ggplot2” package in R, and presented as barplots and boxplots. Subsequently, we analyzed the expression profile of immune checkpoint inhibitor genes in high and low-risk groups. A significant difference was determined by using the Wilcoxon test with p -value < 0.05 . Based on the functional enrichment of gene expression, SubMap module from GenePattern was used to map and merge two datasets with different traits. This module can eliminate the batch effect and predict the possible traits that are not included in the original dataset. To predict the therapeutic response by checkpoint inhibitor blocking, SubMap module was used to map the high and low-risk groups based on the therapeutic information. Prediction of the response to CTLA4 and PD1 inhibitors was particularly studied in high and low-risk NSCLC patients. The

p -value was corrected by Bonferroni to increase the predictive sensitivity. Finally, we used R package “pRRophetic” to predict chemotherapeutic responses of each sample based on the genomics of drug sensitivity in the cancer (GDSC) database. The half-maximal inhibitory concentration (IC50) of each sample was calculated using ridge regression ($p < 0.001$). The Spearman correlation (Cor) between the RS score and IC50 to particular drugs was calculated and the significant correlations were cutoff with $|\text{Cor}| > 0.1$ and $p < 0.001$.

Gene Set Variation Analysis in High and Low-Risk Groups

GSVA analysis was conducted to reveal the differential pathways between high and low-risk patients. Functional enrichment analysis was conducted based on the distinct GIRlncSig patterns extracted from KEGG and MSigDB database using “GSVA” package in R. Significant differential signal pathways between high and low-risk groups were extracted by “limma” package with the threshold for adjusted $p < 0.05$ and $|\log\text{FC}| > 0.1$. Visualized heatmap of KEGG pathway was plotted using “ComplexHeatmap” package in R.

Construction and Validation of the Nomogram Score System

To develop a nomogram score system for NSCLC patients, MCRA was performed to extract the powerful OS predictors. To visualize the results of Cox regression and predict the survival of NSCLC patients, the prognostic nomogram was plotted by using “rms” and “survival” packages in R. First, the Cox proportional hazards regression model was constructed by using the `cph()` function, and the survival (`survival()`) function was used to calculate the survival probability. Finally, the nomogram (`nomogram()`) function was used to create the nomogram, time-dependent ROC, and calibration curves of OS. The predictive performance of the nomogram was validated by calibration curves, ROC, and decision curve analysis (DCA). The area under curves (AUCs) of the nomogram for predicting 1-, 3-, 5-, and 10-year OS of NSCLC patients were plotted.

Statistical Analyses

The chi-squared test, Student’s t -test, Wilcoxon test, and Mann–Whitney U -test were employed to examine the differential variables from datasets or groups. Statistical significance was considered as $p < 0.05$. R (version 3.6.3) was used to perform all statistical analyses, and the results were visualized by corresponding functional packages.

RESULTS

Identification of GIRlncRNAs for NSCLC Patients

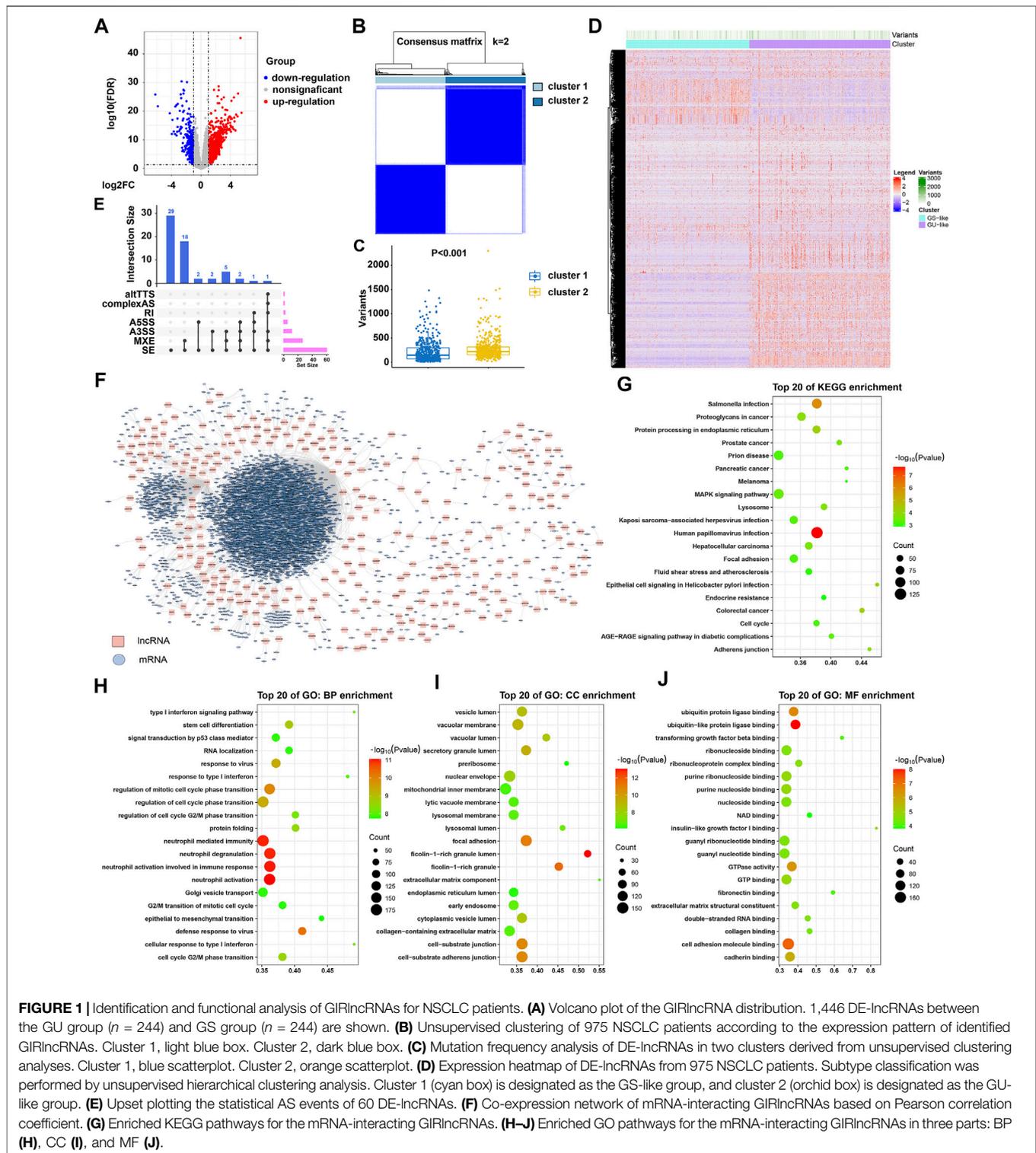
Following TCGA data collection, annotation, and preprocessing, 975 samples were obtained. These samples were subsequently

ranked by their gene mutation counts. The top 25% samples ($n = 244$) with a low mutation frequency were designated as the GS group, and the last 25% samples ($n = 244$) with a high mutation frequency were designated as the GU group (**Supplementary Figure S2; Supplementary Table S1**). A total of 1,446 DE-lncRNAs between these two groups were identified from an expression matrix containing 10,480 lncRNAs ($|\log_2 \text{FC}| > 1$, FDR adjusted $p < 0.05$), among which 1138 lncRNAs were upregulated and 308 lncRNAs were downregulated (**Figure 1A; Supplementary Table S2**). Based on the expression profile of these DE-lncRNAs, all samples were clustered under an unsupervised hierarchical clustering analysis by “ConsumusClusterPlus” package. Two clusters were obtained as shown in the heatmap (**Figure 1B; Supplementary Table S3**). The resulting subtypes were positively correlated with the mutation frequency ($R = 1.22$; $p = 0.0001$) and the mutation frequency in cluster 2 was significantly higher than that in cluster 1 (**Figure 1C**). Therefore, cluster 1 and cluster 2 were named as GS and GU-like subtypes, respectively. As shown by the heatmap of DE-lncRNAs, a significant expression difference was observed between these two subtypes (**Figure 1D**). These results identified 1,446 lncRNAs as candidates for GIRlncRNAs of NSCLC.

Further analyses found that 60 candidate GIRlncRNAs can undergo AS events, of which all DE-lncRNAs contained skipped exon (SE), and 26 DE-lncRNAs contained mutually exclusive exons (MXE) (**Figure 1E; Supplementary Figure S3; Supplementary Table S4**). Co-expression network analyzing the interaction of DE-lncRNA with mRNA resulted in 4912 interaction pairs (**Figure 1F; Supplementary Table S5**), among which 3,802 (77.4%) mRNAs and 334 (6.8%) mRNAs were respectively interacting with lncRNA FENDRR and BANCR. It was also found that 38 and 5 lncRNAs were, respectively, interacting with mRNA AR and TNPO2 (**Supplementary Table S6**). GO enrichment and KEGG pathway analyses showed that these mRNA-interacting GIRlncRNAs were significantly enriched in a nucleoside or ribonucleoside binding, nuclear membranes enveloping, cell-cycle checkpoint, tumorigenesis, and virus infection signaling pathways (**Figures 1G–J; Supplementary Tables S7, S8**). These results suggested that the identified mRNA-interacting GIRlncRNAs may be essential for regulating genomic stability.

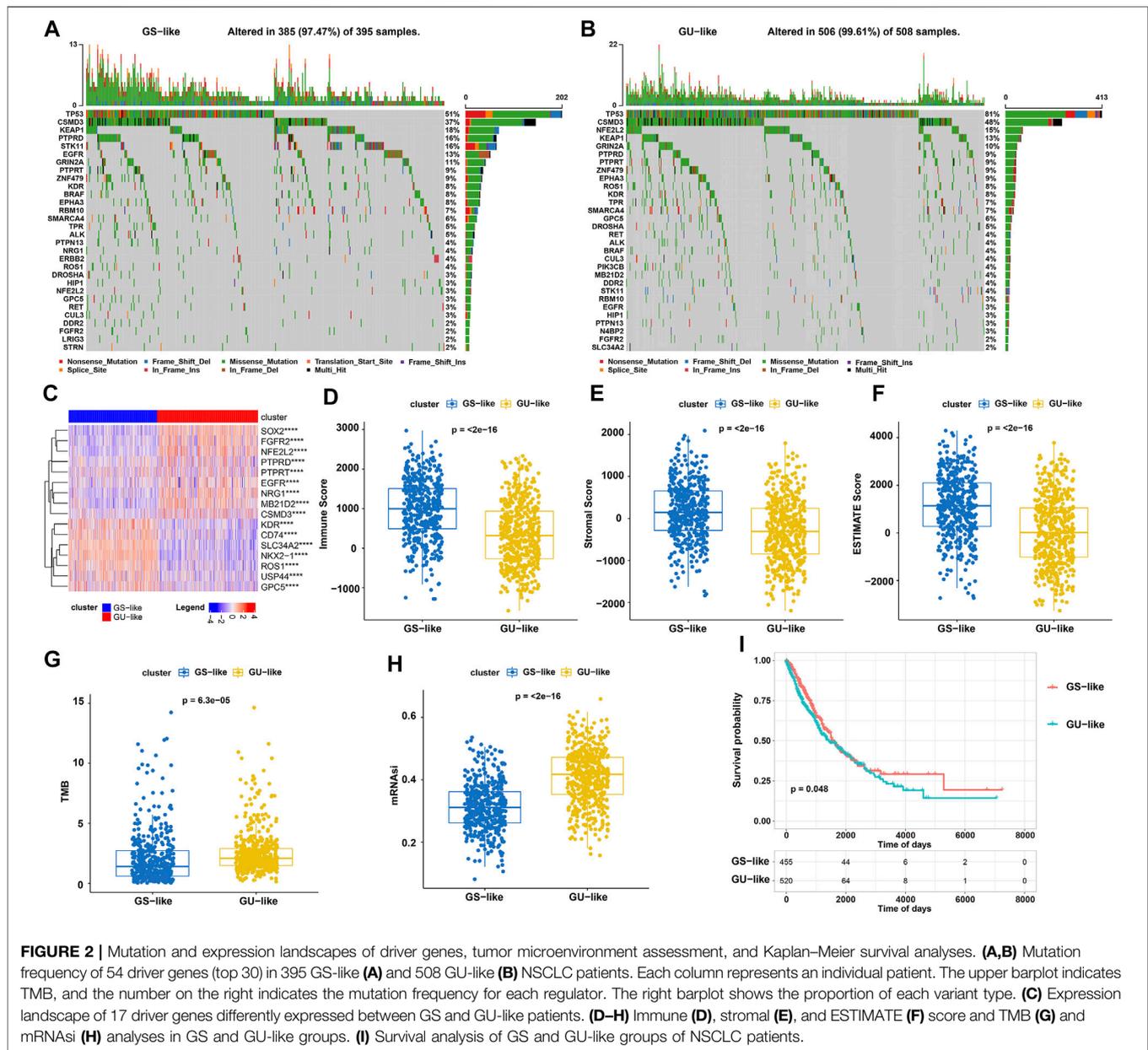
Characteristics of GIRlncRNA-Clustered Molecular Subtypes in NSCLC Patients

Investigating the mutation profile of driver genes showed that GU-like subtype patients exhibited higher mutation frequency than that in GS-like subtype patients. Being the most frequent mutation genes in NSCLC, TP53 and CSMD3 were mutated in 81% and 48% of GU-like subtype patients, respectively. However, these two gene mutations only occurred respectively in 51% and 37% of GS-like subtype patients (**Figures 2A,B**). Further expression analysis of 54 driver genes showed that 16 genes were differentially expressed between GS and GU-like subtypes, among which nine genes were upregulated and seven



genes were downregulated in GU-like patients (Figure 2C). Tumor microenvironment scoring results showed that the immune, stromal and ESTIMATE scores in GU-like subtypes were significantly lower than those in GS-like subtypes (Figures 2D–F). Further analysis of NSCLC tumor characteristics showed

that TMB and mRNA_{si} scores in GU-like subtypes were significantly higher than those in GS-like subtypes (Figures 2G,H). Moreover, Kaplan–Meier survival analysis showed that the survival of GS-like patients was significantly better than that of GU-like patients (Figure 2I). These data indicated that NSCLC



patients with the GU-like molecular subtype have more aggressive tumors than those with GS-like molecular subtypes.

Identification and Evaluation of GIRlncSig for NSCLC Patients

Univariate Cox proportional hazards regression analysis showed that the expressions of 23 GIRlncRNAs were significantly associated with OS of NSCLC patients in our training set ($p < 0.01$ **Figure 3A**; **Supplementary Table S9**). LASSO regression analysis was performed to identify the GIRlncRNAs with more significant associations ($p < 0.05$). When the partial likelihood deviation reached to the minimum (Log Lambda = -4.4), 19 GIRlncRNAs were screened out and used to construct a

risk model for survival prediction (**Figures 3B,C**; **Supplementary Table S10**). To assess the prognostic risk of NSCLC patients, a GIRlncSig was created based on the expression level of 19 GIRlncRNAs and the coefficients from LASSO analysis. GIRlncSig score = $LINC01067^* (-0.1595) + AC012645.1^* (-0.1235) + AL512604.3^* (-0.0989) + AC008278.2^* (-0.0921) + AC089998.1^* (-0.0401) + AC023824.3^* 0.0050 + AC013287.1^* 0.0073 + AP000829.1^* 0.0159 + LINC01611^* 0.0177 + AC097451.1^* 0.0197 + AC025419.1^* 0.0270 + AC079949.2^* 0.0369 + LINC01600 + 0.0474 + AC004862.1^* 0.0518 + AC021594.1^* 0.0623 + (MYRF-AS1)^* 0.0851 + LINC02434^* 0.0858 + LINC02412^* 0.1049 + LINC00337^* 0.1143$. In the formula of GIRlncSig, 14 GIRlncRNAs (AC023824.3, AC013287.1, AP000829.1,

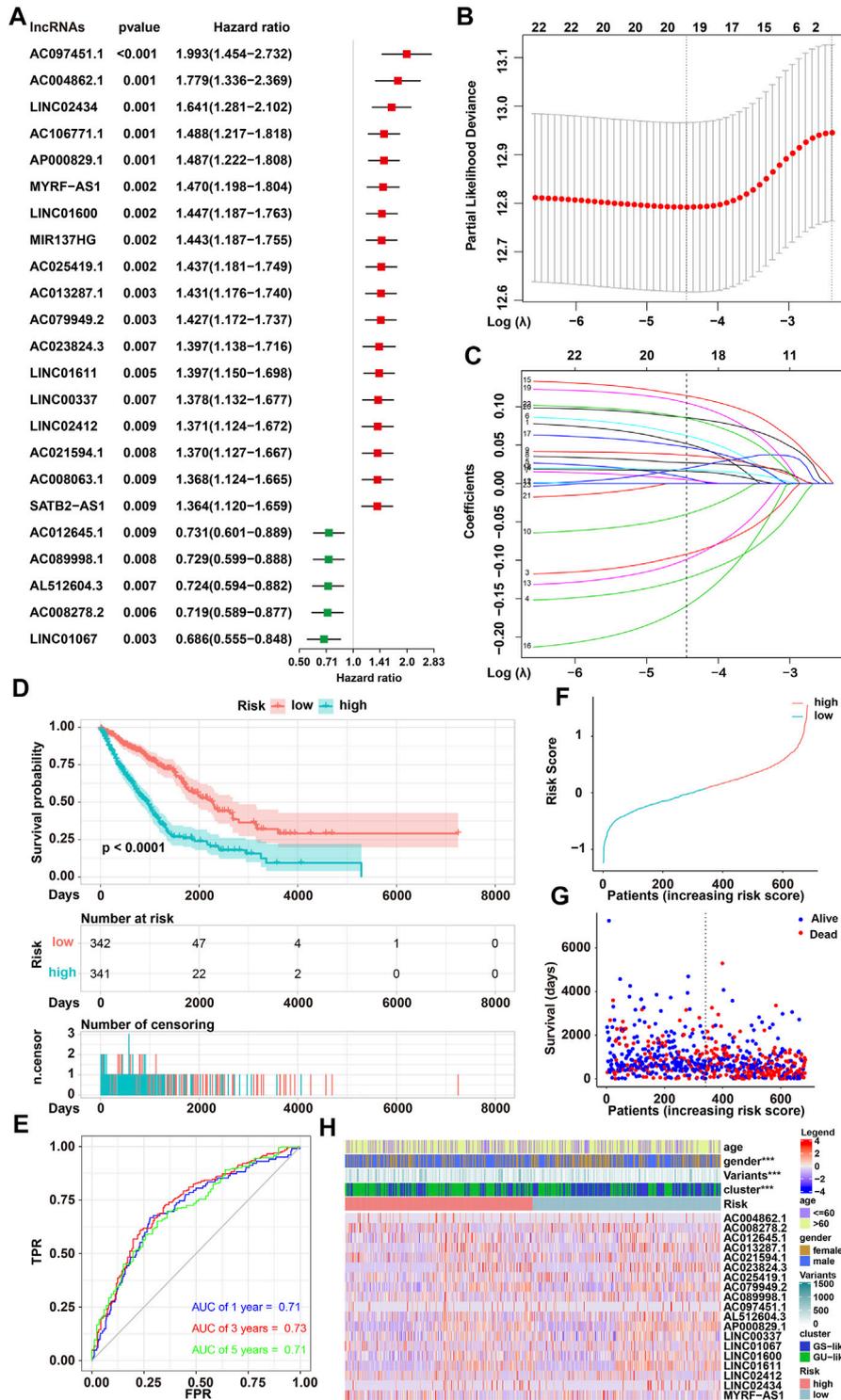


FIGURE 3 | Identification of the GIrIncSig and its predictive performance in a training set. **(A)** Forest plot of OS-associated GIrIncRNAs based on the univariate Cox proportional hazards regression analysis. Five GIrIncRNAs acted as protective risk factors for patients' survival (green), while 18 GIrIncRNAs acted as prognostic risk factors (red). **(B)** Distribution plot of partial likelihood deviation based on LASSO regression analysis. Nineteen GIrIncRNAs were selected when log lambda was equal to -4.4 (the minimum). **(C)** Distribution plot of LASSO coefficient (log lambda = -4.4). **(D)** Kaplan–Meier survival curves of low and high-risk NSCLC patients predicted by the GIrIncSig (log-rank test, $p < 0.0001$). **(E)** ROC of NSCLC patients at 1, 3, or 5 years predicted by GIrIncSig. **(F,G)** Risk distribution **(F)** and survival status **(G)** of NSCLC patients. **(H)** Expression heatmap of selected GIrIncRNAs.

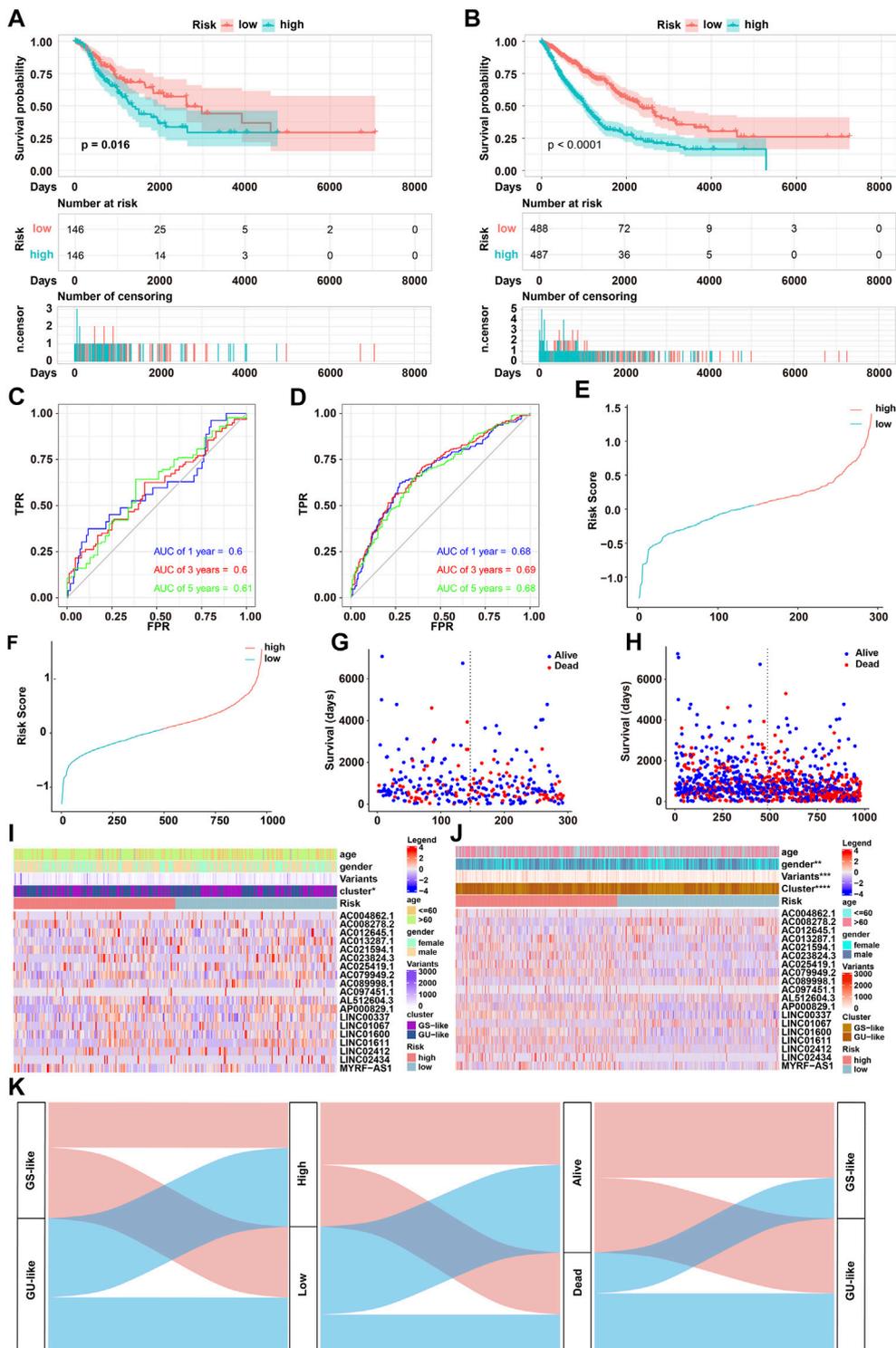


FIGURE 4 | Performance evaluation of GIrIncSig in testing and TCGA sets. **(A,B)** Kaplan–Meier survival curves of GIrIncSig-predicted low- or high-risk NSCLC patients in testing **(A)**, log-rank test; $p < 0.05$] and TCGA **(B)**, log-rank test; $p < 0.0001$] sets. **(C,D)** ROC of 1-, 3-, or 5-year survival predicted by GIrIncSig in testing **(C)** and TCGA **(D)** sets. **(E–J)** Risk distribution, survival status of patients, and the expression profile of GIrIncSig in testing **(E,G,I)** and TCGA **(F,H,J)** sets. **(K)** Sankey diagram for the distribution of NSCLC patients in GS or GU-like groups, low or high-risk groups, and dead or alive groups.

LINC01611, AC097451.1, AC025419.1, AC079949.2, LINC01600, AC004862.1, AC021594.1, MYRF-AS1, LINC02434, LINC02412, and LINC00337) with positive coefficients were designated as risk factors. In contrast, five lncRNAs (LINC01067, AC012645.1, AL512604.3, AC008278.2, and AC089998.1) with negative coefficients were designated as protective factors. Either upregulated expression of prognostic GIRlncRNAs or downregulated expression of protective GIRlncRNAs was significantly related to decreased OS of NSCLC patients. Based on the GIRlncSig score (cutoff = 0.075), 341 patients with a high score were classified into a high-risk group, and 342 patients with a low score were classified into a low-risk group (**Supplementary Table S11**). Kaplan–Meier survival analysis showed that the survival of low-risk NSCLC patients was significantly better than that in high-risk patients ($p < 0.0001$, log-rank test; **Figure 3D**). The AUCs of time-dependent ROC in the training set were 0.71, 0.73, 0.71 for 1-, 3-, 5-year survival, respectively, as predicted by GIRlncSig (**Figure 3E**). Finally, the risk distribution and survival status of NSCLC patients together with the expression heatmap of 19 GIRlncRNAs in the training set were plotted (**Figures 3F–H**). The results showed that high-risk-scored patients with such a GIRlncSig were mainly derived from the GU-like group and exhibited shorter survival. All these results strongly supported the utility and effectiveness of our GIRlncSig in predicting NSCLC prognosis.

To evaluate the robustness of our GIRlncSig, its prognostic performance in two independent data sets was further tested. A total of 392 NSCLC patients from the testing set and 975 NSCLC patients from the TCGA set were classified into high- and low-risk groups based on the GIRlncRNA risk score (**Supplementary Table S11**). Kaplan–Meier survival curves in the testing set ($p < 0.016$, **Figure 4A**) and entire TCGA set ($p < 0.0001$, **Figure 4B**) showed that patients from the low-risk group had better survival outcomes than patients from the high-risk group. The AUCs of time-dependent ROC in the testing set were 0.6, 0.6, and 0.61 for the 1-, 3-, and 5-year survival predicted by the GIRlncSig, respectively (**Figure 4C**). Similar results were obtained in TCGA set, where the AUCs of ROC were overall approximately 0.7 (**Figure 4D**). Furthermore, risk distribution, and survival of patients together with GIRlncSig expression heatmap in both sets showed that high-risk scored patients were mainly derived from the GU-like group and exhibited shorter survival (**Figures 4E–J**). Additionally, the Sankey diagram also showed that high-risk patients accounted for a higher proportion of mortality in the GU-like group, while these patients accounted for a less proportion of mortality in the GS-like group (**Figure 4K**). Together, these results further support that our GIRlncSig can predict the prognosis of NSCLC patients.

Correlation of GIRlncSig With the Aggressiveness of NSCLC

Since GIRlncSig possessed a robust prognostic performance, the correlation analysis between identified GIRlncRNAs and differentially expressed driver genes was further performed. Three positively correlated regulatory pairs (AC008278.2 and PTPRT, AP000829.1 and MB21D2, LINC01600 and MB21D2) were

screened out by Pearson correlation analysis ($R > 0.3$ and $p < 0.05$) (**Figures 5A–C**; **Supplementary Table S12**). The expression patterns of these correlated pairs were consistent in high- and low-risk groups (**Figures 5D–F**). To determine the relationship of GIRlncSig with tumor microenvironment characteristics, Pearson correlation coefficients between GIRlncSig risk score and immune, or stromal, or ESTIMATES, or mRNAsi, or TMB scores were calculated separately. The results of correlation analysis showed that GIRlncSig-based risk scores were positively correlated with all characteristics of the tumor microenvironment (**Figures 5G–K**). Notably, the strongest correlation of the risk score with TMB was observed ($R = 0.135$; **Figure 5K**). Further correlation analyses between the GIRlncSig and infiltrating immune cells were performed. Although the resulted NSCLC-infiltrated immune cell types were different depending on algorithms, similar trends of cell distribution were as follows: high infiltrating level of B cell (especially memory B cell) and T cell (especially CD4⁺ and CD8⁺ memory T cells) populations were positively associated with the expression of five protective lncRNAs. By contrast, infiltrating B and T cells presented an opposite association trend for fourteen risk lncRNAs (**Supplementary Figures S4–S10**). High infiltrating levels of macrophages (especially M1 and M2), neutrophils, monocytes, and myeloid dendritic cells were positively associated with the expression of protective lncRNAs but negatively associated with the expression of risk lncRNAs (**Supplementary Figures S4–S10**). These results suggested that GIRlncSig was significantly correlated with the malignancy of NSCLC.

Risk Stratification of NSCLC Patients With GIRlncSig Score and Clinical Variables

To validate the stability of our score model, a risk stratification analysis was conducted to determine the prognostic performance of GIRlncSig. NSCLC patients were first grouped based on their clinical parameters, then stratified into subgroups by GIRlncSig-derived risk score. Kaplan–Meier survival analyses showed that patients with low-risk scores had better survival outcomes than those with high-risk scores in all stratified subgroups ($p < 0.05$; **Figures 6A–H**). The distribution of risk scores for all stratified subgroups, including age, gender, pathologic M, pathologic N, pathologic T, tumor stage, and GS/GU-like groups, was further determined. No difference in the distribution of risk score between young and old NSCLC patients was observed ($p < 0.05$; **Figure 6I**). However, the risk score distribution of male patients was significantly higher than that of female patients ($p < 0.05$; **Figure 6J**), and the risk score distribution of the GU-like group was significantly higher than that in the GS-like group ($p < 0.05$; **Figure 6O**). For the groups of pathologic N, T, and tumor stage but not pathologic M, the risk score for the primary stage was always low. Interestingly, higher risk scores were frequently observed in patients with advanced stage ($p < 0.05$; **Figures 6K–N**). These results highlight the stability of our GIRlncSig-based risk score model.

Therapeutic Evaluation of NSCLC Patients by the GIRlncSig Score

The abundance of tumor-infiltrating immune cells and the expression profile of immune checkpoint genes have a strong impact on tumor treatment, so we carried out immune cell

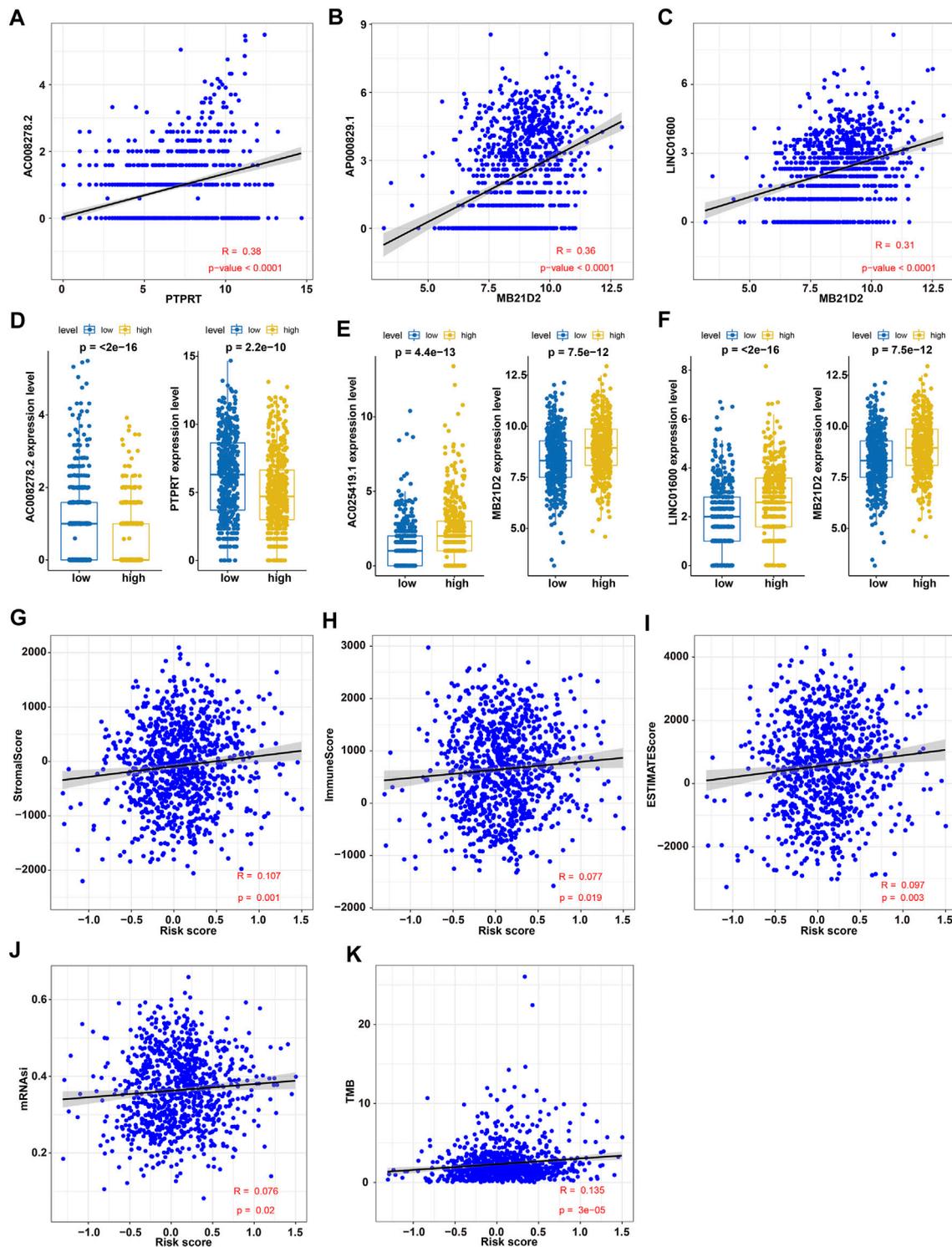


FIGURE 5 | Correlation analyses between GIRlncSig and malignant characteristics of NSCLC. **(A–C)** Pearson correlation analysis of GIRlncRNA-driver gene regulatory pairs: AC008278.2 and PTPRT **(A)**, AP000829.1 and MB21D2 **(B)**, LINC01600 and MB21D2 **(C)**; **(D–F)** Differential expression analysis of correlated regulatory pairs in different risk groups: AC008278.2 and PTPRT **(D)**, AP000829.1 and MB21D2 **(E)**, and LINC01600 and MB21D2 **(F)**. **(G–K)** Pearson correlation analysis between the risk score and immune score **(G)**, stromal score **(H)**, ESTIMATE score **(I)**, mRNAasi score **(J)**, or TMB score **(K)** in NSCLC patients.

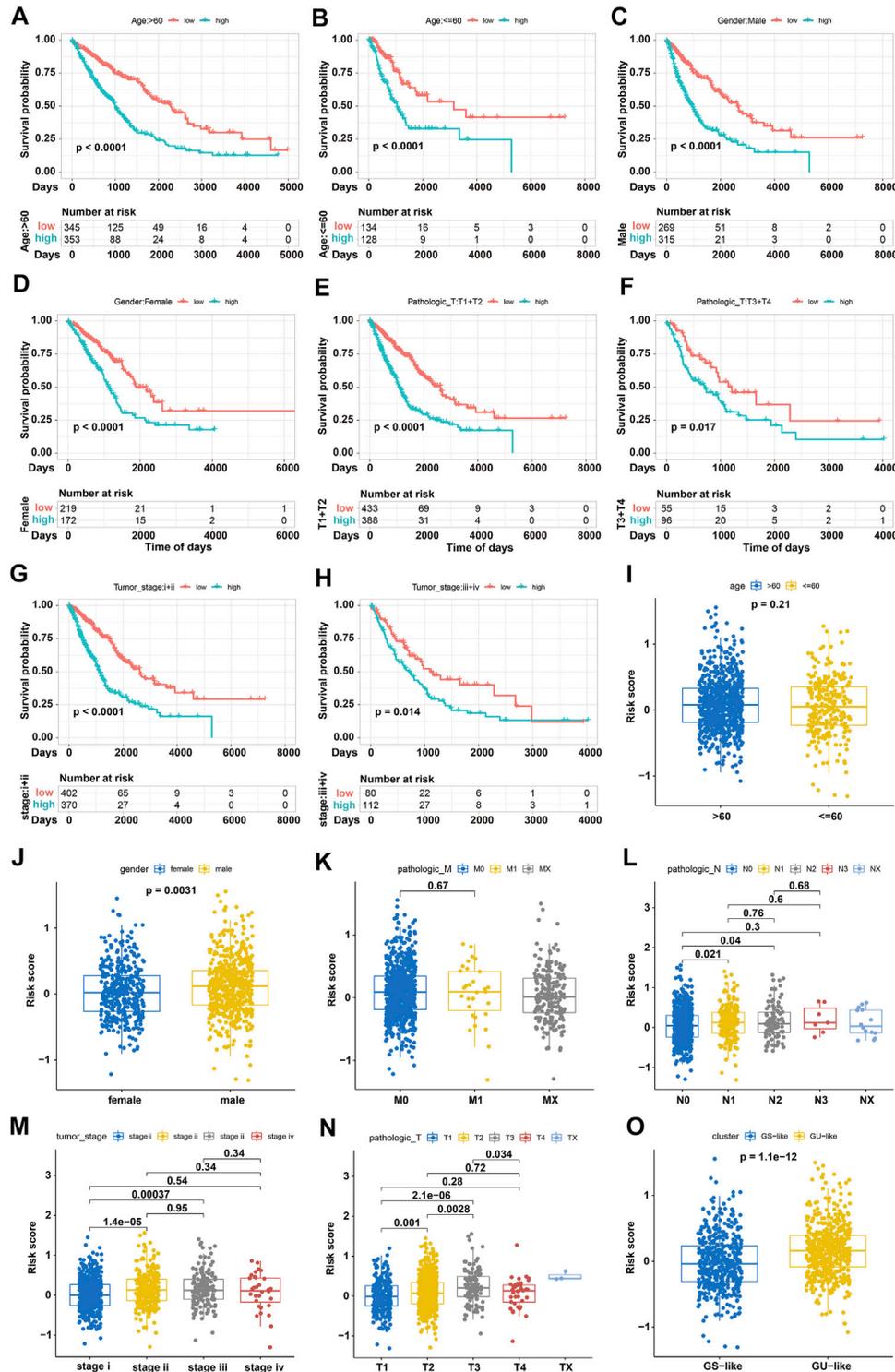
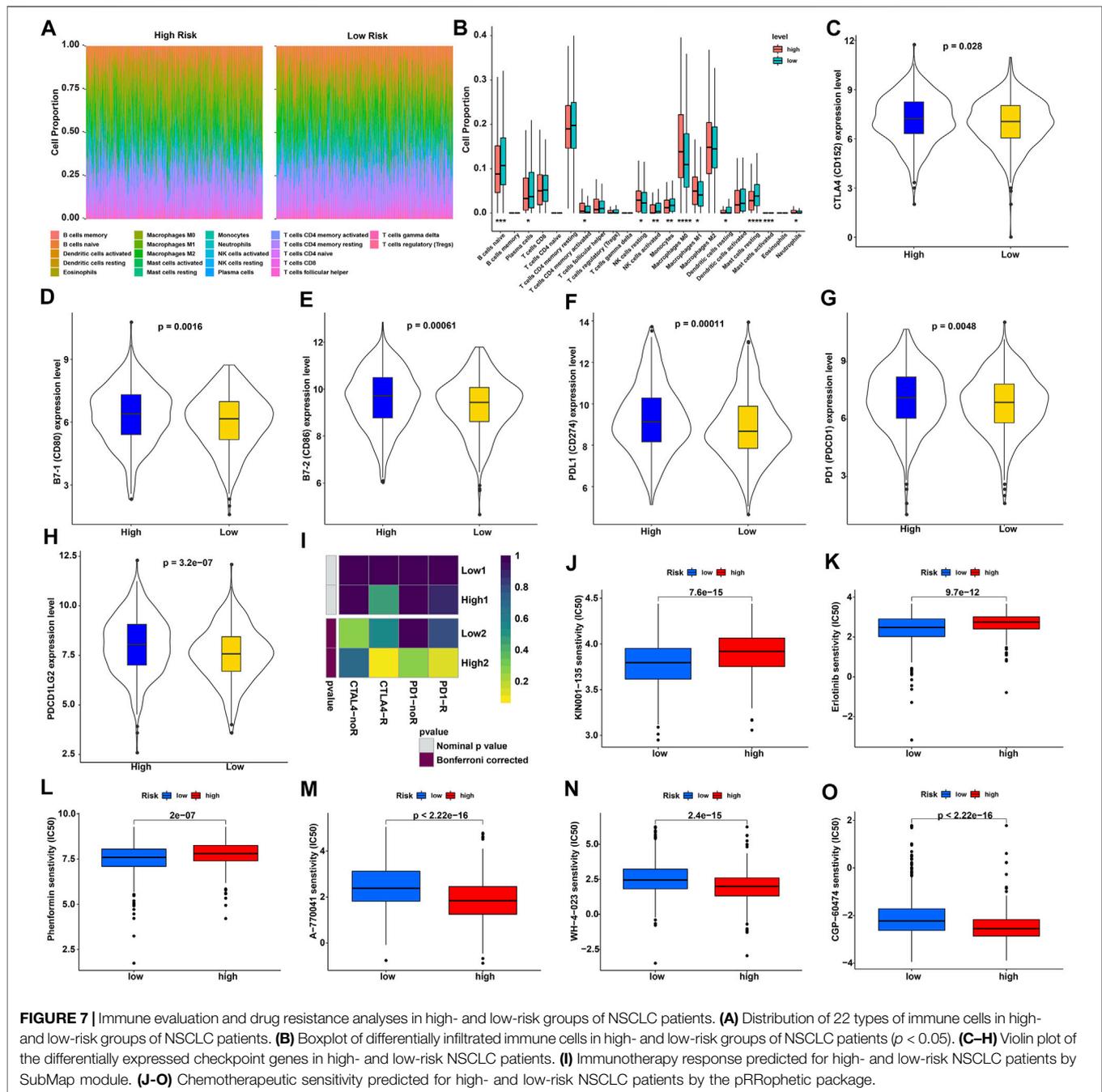


FIGURE 6 | Risk stratification analyses based on GIrlncSig score and clinical variables. **(A–H)** Kaplan–Meier survival curves of high and low-risk subgroups for old **(A)**, young **(B)**, male **(C)**, female **(D)**, T1–2 stage **(E)**, T3–4 stage **(F)**, stage I–II **(G)**, and stage III–IV **(H)** patients. **(I–N)** Boxplots showing GIrlncSig-derived risk scores stratified by age **(I)**, gender **(J)**, and pathologic M **(K)**, N **(L)**, T **(M)**, and tumor stage **(N)** in NSCLC patients. **(O)** Boxplot of GIrlncSig-derived risk scores in GS- and GU-like NSCLC groups.



infiltration analysis with the CIBERSORT algorithm to evaluate our GIRIncSig. Abundance ratios and differential boxplots of 22 types of immune cells in high and low-risk NSCLC patients were plotted (Figure 7A; Supplementary Table S13). The abundance ratios of infiltrated M0, M1 macrophages, and resting NK cells in the high-risk score group were significantly higher than those in low-risk score group. However, naive B cells, plasma cells, monocytes, and resting mast cells in the high-risk score group were remarkably lower than those in the low-risk score group (Figure 7B). Subsequently, the expression profile of immune checkpoint genes, CTLA4 (CD152), B7-1 (CD80), B7-2

(CD86), PDL1 (CD274), PD1 (PDCD1) and PDL2 (PDCD1LG2), in high and low-risk groups was further analyzed (Figures 7C–H). The violin plots showed a significant difference in the expression levels of CTLA4, CD80, CD86, CD274, PDCD1, and PDCD1LG2 between low and high-risk groups (Figures 7C–H). Therefore, the SubMap module in GenePattern database was further employed to predict the risk score effect on immunotherapy of NSCLC patients. Results of the corrected Bonferroni analysis suggested that the patients in the high-risk group were slightly more sensitive to CTLA4 and PD1 inhibitors than those in the low-risk group ($p = 0.0619$ and 0.0739 ;

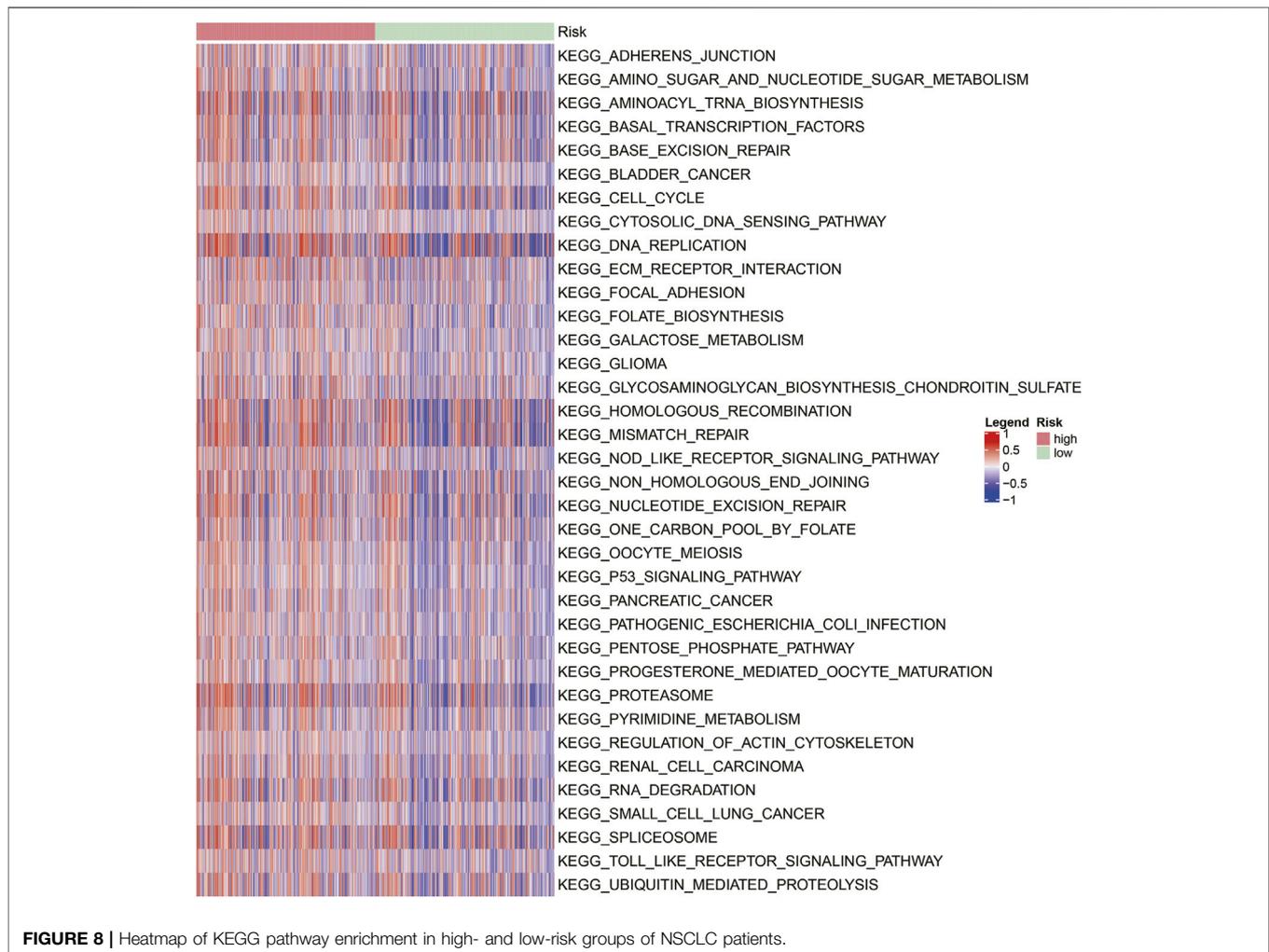


Figure 7I). We also evaluated the chemotherapeutics sensitivity in different risk groups predicted by pRRophetic package. It was found that a total of 76 drugs, including four resistant drugs and 72 sensitive drugs, were correlated with the RS scores (**Supplementary Figure S11**). The predicted results showed that NSCLC patients in the high-risk group were more resistant to KIN001-135, erlotinib and phenformin than those in the low-risk group. However, these high-risk NSCLC patients were more sensitive to A-770041, WH-4-023, and CGP-60474 (**Figures 7J–O**). Overall, these results suggested that GIRlncSig could be used for the evaluation of immune cell distribution and immunotherapy response in NSCLC patients.

Genomic Instability-Related Signal Pathways Were Enriched in High-Risk Patients

To explore the biological function associated with the GIRlncSig, functional enrichment analysis was conducted for high and low-risk groups of NSCLC patients using “GSVA” package. MSigDB database-based KEGG analysis revealed that 32 differentially

enriched items were significantly enriched in the high-risk scored group (**Figure 8**). Notably, two types of signaling pathways were markedly enriched in the high-risk scored group. One was the nucleic acid metabolic pathway including pyrimidine metabolism, folate biosynthesis, DNA replication, and RNA degradation. The other was the DNA damage repair pathway including mismatch repair, base excision repair, nucleotide excision repair, non-homologous end joining, and homologous recombination. Notably, these two types of pathways were strongly associated with genomic stability.

Independent Prognostic Evaluation and Nomogram Construction Based on the Risk Score and Clinical Variables

To verify whether GIRlncSig was an independent prognostic factor, UCRA was conducted on variables including clinical variables and GIRlncSig-based risk score. Then, MCRA was used to evaluate the prognosis of all included variables. The UCRA and MCRA results showed that the GIRlncSig-based risk score and tumor stage exhibited good prognostic performance

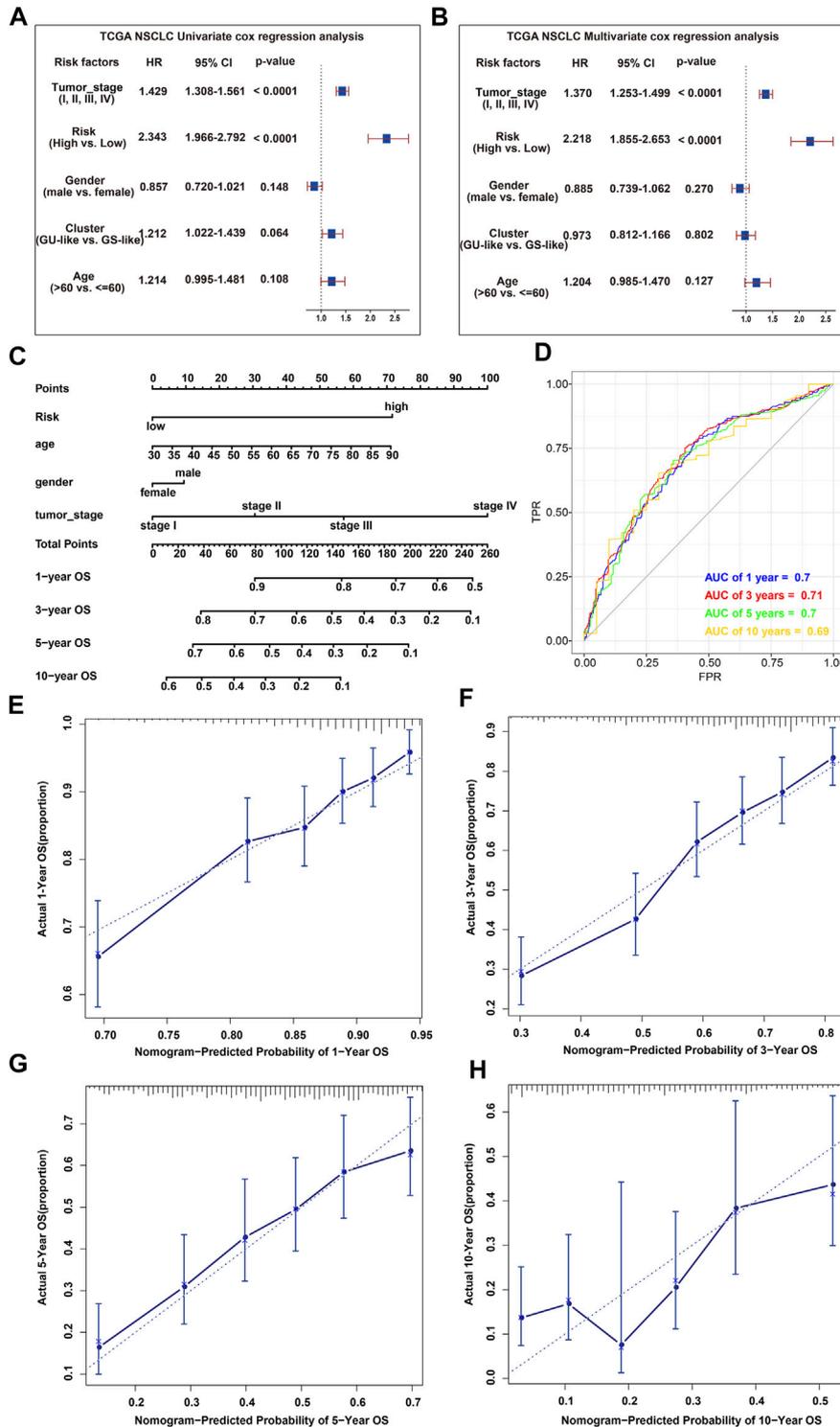


FIGURE 9 | Independent prognostic evaluation and nomogram construction. **(A,B)** UCRA **(A)** and MCRA **(B)** based on GIIncSig-based risk scores together with clinical variables; **(C)** MCRA-developed nomogram for predicting 1-, 3-, 5-, and 10-year survival of NSCLC patients; **(D)** MCRA-developed ROC for predicting 1-, 3-, 5-, and 10-year survival of NSCLC patients; **(E–H)** Calibration curves for predicting 1- **(E)**, 3- **(F)**, 5- **(G)**, and 10- **(H)** year survival of NSCLC patients.

($p < 0.0001$; **Figures 9A,B; Supplementary Table S14**). Other variables, such as cluster, age, and gender, had no significant correlation with the OS of NSCLC patients ($p > 0.05$). These results suggested that the prognostic value of GIRlncSig was independent of other clinical variables. To test its clinical utility, a statistical nomogram was created by integrating GIRlncSig-based risk score with TCGA clinical information (age, gender, and tumor stage) (**Figure 9C**). The C-index of nomogram was 0.67, and the AUCs of ROC for 1-, 3-, 5-, and 10-year survival predictions were 0.69, 0.71, 0.70, and 0.69, respectively (**Figure 9D**). The calibration plots of 1-, 3-, 5-, and 10-year OS showed good agreements between the actual survival rate and the nomogram-predicted survival rate (**Figures 9E–H**). Therefore, these data suggested that the nomogram has a good prediction performance and could provide clues for clinical diagnosis of NSCLC.

DISCUSSION

Genomic instability is an evolving hallmark of most cancers (Hanahan, 2022). It is also a major driver of carcinogenesis, drug sensitivity, tumor-microenvironment-shaping, and immune contexture in NSCLC (Raynaud et al., 2018; Skoulidis and Heymach, 2019). lncRNAs play a critical role in maintaining genomic instability (Nair et al., 2020; Jianfeng et al., 2021). Increasing evidence has revealed the prognostic significance of GIRlncRNAs for cancers (Fang et al., 2021; Liang et al., 2021; Maimaiti et al., 2021; Yan et al., 2021). Since NSCLC possesses a poor survival prognosis due to limited diagnosis and treatment (Zappa and Mousa, 2016; Li C. et al., 2019), we developed a prognostic GIRlncSig to support the clinical stratification and treatment decision for NSCLC patients.

In this study, we screened out 1446 GIRlncRNAs for NSCLC by a somatic mutation burden hypothesis-derived computational frame. Functional analysis revealed that these lncRNAs were mainly enriched in nucleoside or ribonucleoside metabolism, cell-cycle checkpoint, nuclear membrane enveloping, and tumorigenesis, which are involved in maintaining genomic instability (Sieber et al., 2003; Deng, 2006; Aird and Zhang, 2015; Lim et al., 2016). TP53 and CSMD3 were the two most frequently mutated genes in NSCLC (Liu et al., 2012), and their mutation status was closely associated with high TMB causing genomic instability and poor clinical prognosis (Zhang et al., 2017; Bernard et al., 2020; Lu et al., 2021; Wen et al., 2021). We further conducted hierarchical clustering analysis and differential analysis of mutation counts and found that GIRlncRNA-clustering GU-like patients were burdened with a higher TMB than GS-like patients. The mutation frequencies of TP53 and CSMD3 genes in GS-like subtype patients were expectedly higher than those in GU-like subtype patients. The survival of GS-like subtype patients was significantly better than that of GU-like subtype patients. Furthermore, we constructed a GIRlncSig encompassing 19 lncRNAs with robust performances, which could predict prognosis independently of other clinicopathological variables and data sets. Among the 19 GIRlncRNA signature, 14 lncRNAs (AC023824.3, AC013287.1, AP000829.1, LINC01611, AC097451.1, AC025419.1, AC079949.2, LINC01600, AC004862.1,

AC021594.1, MYRF-AS1, LINC02434, LINC02412, and LINC00337) were risk factors for prognosis, while the other five lncRNAs (LINC01067, AC012645.1, AL512604.3, AC008278.2, and AC089998.1) were protective factors for survival of NSCLC patients. To the best of our knowledge, most lncRNAs we identified here are novel GIRlncRNAs for NSCLC, while some GIRlncRNAs were already reported in lung adenocarcinoma (Geng et al., 2021; Peng et al., 2021; Wu G. et al., 2021). Notably, lncRNA AC023824, AC025419.1, AC079949.2, LINC02412, and LINC00337 were verified as risk factors associated with the OS of LUAD patients (Li R. et al., 2019; Song et al., 2020; Shao et al., 2021; Wu G. et al., 2021; Wu Y. et al., 2021). LINC01600 and LINC02434 were reported as predictors for the prognosis of PCa and HNSCC patients, respectively (Xu et al., 2020; Jiang et al., 2021). Importantly, we performed ROC and calibration analyses to evaluate the GIRlncSig-based risk score and found that it possessed an intact performance with good agreement between the actual survival and predicted survival in 10 years. In contrast, Wang and Geng's prognostic GIRlncSig only displayed a decreased value of AUCs in 3 years. Therefore, our novel GIRlncSig could provide robust clues for clinical diagnosis and stratification of NSCLC.

Mutations in driver genes are crucial to promoting tumorigenesis and development. NSCLC with positive driver genes possesses high mortality and metastasis risk (Wu Y. et al., 2021; Yuan et al., 2022). We found here that nine driver genes (SOX2, FGFR2, NFE2L2, PTPRD, PTPRT, EGFR, NRG1, MB21D2, and CSMD3) were upregulated. Notably, the expression levels of CSMD3, NFE2L2, and MB21D2 were substantially higher in GU-like samples than those in GS-like samples. NFE2L2 mutation, a major molecular driver of clinical radio resistance (Binkley et al., 2020), was more frequently found in advanced patients to cause a worse prognosis than in patients carrying the wild-type genotype (Sasaki et al., 2010). MB21D2, a key enzyme involved in the cGAS/STING signaling pathway, is also frequently mutated in NSCLC and HNSCC to promote tumor progression (Campbell et al., 2016; Gracilla et al., 2020). The high risk lncRNAs AP000829.1 and LINC01600 from GIRlncSig were positively correlated with MB21D2 expression, which was remarkably upregulated in the high-risk group. Another high-risk lncRNA, AP000829.1, was negatively correlated with NKX2-1. In fact, upregulated AP000829.1 was always accompanied by downregulated NKX2-1 in the high-risk group. NKX2-1 may control lung cancer progression through the induction of DUSP6, an ERK phosphatase, to decrease ERK activity (Ingram et al., 2022). The protective lncRNA, AC008278.2, was positively correlated with the driver gene PTPRT, and this pair was downregulated in the high-risk group. Because PTPRT is an endogenous inhibitor of STAT3 (Sen et al., 2020), loss-of-function mutations in PTPRT resulted in STAT3 hyperactivation to promote the malignancy of NSCLC (Wang et al., 2021). Moreover, the functional genes markedly enriched in the high-risk group mainly involve mismatch repair, base excision repair, nucleotide excision repair, non-homologous end joining, and homologous recombination. These pathways were evidenced to be strongly associated with genomic stability (Majidinia and Yousefi, 2017). Together, these findings suggested that the lncRNAs of our

GIRlncSig are remarkably modulated by the NSCLC driver genes, and that their functions are highly consistent with important biological behaviors.

The contextures of the tumor microenvironment, like infiltrated immune cells, stromal cells, cancer stem cells, and TMB, critically determine the progression of cancer (Whiteside, 2008). Hence, pursuing them can help to predict clinical outcomes, guide early diagnosis and improve the therapeutic response (Binnewies et al., 2018). We found here that the novel GIRlncSig could reflect the characteristics of the tumor microenvironment, and GIRlncSig-predicted high-risk patients exhibited features of malignancy with high levels of TMB and mRNA_{si}, as well as immune, stromal, and ESTIMATE scores. These results accord with the fact that genome instability contributes to neoplasia and metastasis (Bakhoun et al., 2018; Nguyen et al., 2022). Moreover, we found that NSCLC patients from high-risk groups were more sensitive to CTLA4 inhibitors than those from the low-risk group. Collectively, our GIRlncSig may guide the diagnosis and improve the clinical outcome of NSCLC by selecting a subgroup of patients that are more sensitive to this type of immunotherapy.

Two limitations are associated with this study. First, as appropriate GEO datasets were not found, some potential lncRNAs may have been excluded in our GIRlncSig. Hence, more external dataset validation is needed in future studies. Second, the biological functions of twelve newly identified GIRlncRNAs (AC013287.1, AP000829.1, LINC01611, AC097451.1, AC004862.1, AC021594.1, MYRF-AS1, LINC01067, AC012645.1, AL512604.3, AC008278.2, and AC089998.1) are not known yet. Future investigations should elucidate their functions, both *in vitro* and *in vivo*.

CONCLUSION

In summary, we constructed a GIRlncSig consisting of 19 GIRlncRNAs. This signature could predict the clinical outcome of NSCLC patients independently of other variables. Moreover, our GIRlncSig could dissect the contextures of the tumor microenvironment and driver genes to guide the diagnosis for improved stratification and individualized treatment of NSCLC patients.

REFERENCES

- Aird, K. M., and Zhang, R. (2015). Nucleotide Metabolism, Oncogene-Induced Senescence and Cancer. *Cancer Lett.* 356 (2 Pt A), 204–210. doi:10.1016/j.canlet.2014.01.017
- Bakhoun, S. F., Ngo, B., Laughney, A. M., Cavallo, J. A., Murphy, C. J., Ly, P., et al. (2018). Chromosomal Instability Drives Metastasis through a Cytosolic DNA Response. *Nature* 553 (7689), 467–472. doi:10.1038/nature25432
- Bao, S., Zhao, H., Yuan, J., Fan, D., Zhang, Z., Su, J., et al. (2020). Computational Identification of Mutator-Derived lncRNA Signatures of Genome Instability for Improving the Clinical Outcome of Cancers: a Case Study in Breast Cancer. *Brief. Bioinform* 21 (5), 1742–1755. doi:10.1093/bib/bbz118
- Bastians, H. (2015). Causes of Chromosomal Instability. *Recent Results Cancer Res.* 200, 95–113. doi:10.1007/978-3-319-20291-4_5

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/**Supplementary Material**.

AUTHOR CONTRIBUTIONS

SZ and QZ initiated the idea, designed the experiments, supervised the research, and wrote the manuscript. QZ, XL, and ZC collected, analyzed, and generated data and figures. All authors have read, edited, and approved the final version of this manuscript.

FUNDING

This work was supported by the National Natural Science Foundation of China (Nos. 81373318 and 82073233), the National Translational Science Center for Molecular Medicine Fund (F1034361), the Chinese Pharmaceutical Association-Yiling Pharmaceutical Innovation Fund (CPAYLJ202003), the Key Research Fund of Tianjin Project and Team (XB202010), the Key Research and Development Program of Tianjin (20YFZCSY00450), and the Tianjin sci-tech commissioner Fund (21YDTPJC00220).

ACKNOWLEDGMENTS

The authors greatly thank Dr. Qing Zhang (Tianjin Medical University, China) for discussion and helpful suggestion. They also thank Prof. Mattias Belting (Lund University, Sweden) for linguistic advice.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fphar.2022.937531/full#supplementary-material>

- Bernard, E., Nannya, Y., Hasserjian, R. P., Devlin, S. M., Tuechler, H., Medina-Martinez, J. S., et al. (2020). Implications of TP53 Allelic State for Genome Stability, Clinical Presentation and Outcomes in Myelodysplastic Syndromes. *Nat. Med.* 26 (10), 1549–1556. doi:10.1038/s41591-020-1008-z
- Binkley, M. S., Jeon, Y. J., Nesselbush, M., Moding, E. J., Nabet, B. Y., Almanza, D., et al. (2020). KEAP1/NFE2L2 Mutations Predict Lung Cancer Radiation Resistance that Can Be Targeted by Glutaminase Inhibition. *Cancer Discov.* 10 (12), 1826–1841. doi:10.1158/2159-8290.CD-20-0282
- Binnewies, M., Roberts, E. W., Kersten, K., Chan, V., Fearon, D. F., Merad, M., et al. (2018). Understanding the Tumor Immune Microenvironment (TIME) for Effective Therapy. *Nat. Med.* 24 (5), 541–550. doi:10.1038/s41591-018-0014-x
- Campbell, J. D., Alexandrov, A., Kim, J., Wala, J., Berger, A. H., Pedamallu, C. S., et al. (2016). Distinct Patterns of Somatic Genome Alterations in Lung Adenocarcinomas and Squamous Cell Carcinomas. *Nat. Genet.* 48 (6), 607–616. doi:10.1038/ng.3564

- Chen, J. W., and Dhahbi, J. (2021). Lung Adenocarcinoma and Lung Squamous Cell Carcinoma Cancer Classification, Biomarker Identification, and Gene Expression Analysis Using Overlapping Feature Selection Methods. *Sci. Rep.* 11 (1), 13323. doi:10.1038/s41598-021-92725-8
- Chen, B., Khodadoust, M. S., Liu, C. L., Newman, A. M., and Alizadeh, A. A. (2018). Profiling Tumor Infiltrating Immune Cells with CIBERSORT. *Methods Mol. Biol.* 1711, 243–259. doi:10.1007/978-1-4939-7493-1_12
- Chin, C. H., Chen, S. H., Wu, H. H., Ho, C. W., Ko, M. T., and Lin, C. Y. (2014). cytoHubba: Identifying Hub Objects and Sub-networks from Complex Interactome. *BMC Syst. Biol.* 8 Suppl 4, S11. doi:10.1186/1752-0509-8-S4-S11
- Deng, C. X. (2006). BRCA1: Cell Cycle Checkpoint, Genetic Instability, DNA Damage Response and Cancer Evolution. *Nucleic Acids Res.* 34 (5), 1416–1426. doi:10.1093/nar/gkl010
- Fang, X., Liu, X., Lu, L., and Liu, G. (2021). Identification of a Somatic Mutation-Derived Long Non-Coding RNA Signatures of Genomic Instability in Renal Cell Carcinoma. *Front. Oncol.* 11, 728181. doi:10.3389/fonc.2021.728181
- Feng, J. X., and Riddle, N. C. (2020). Epigenetics and Genome Stability. *Mamm. Genome* 31 (5–6), 181–195. doi:10.1007/s00335-020-09836-2
- Gao, H., Song, X., Kang, T., Yan, B., Feng, L., Gao, L., et al. (2017). Long Noncoding RNA CRNDE Functions as a Competing Endogenous RNA to Promote Metastasis and Oxaliplatin Resistance by Sponging miR-136 in Colorectal Cancer. *Oncotargets Ther.* 10, 205–216. doi:10.2147/OTT.S116178
- Geng, W., Lv, Z., Fan, J., Xu, J., Mao, K., Yin, Z., et al. (2021). Identification of the Prognostic Significance of Somatic Mutation-Derived lncRNA Signatures of Genomic Instability in Lung Adenocarcinoma. *Front. Cell Dev. Biol.* 9, 657667. doi:10.3389/fcell.2021.657667
- Gracilla, D. E., Korla, P. K., Lai, M. T., Chiang, A. J., Liou, W. S., and Sheu, J. J. (2020). Overexpression of Wild Type or a Q311E Mutant MB21D2 Promotes a Pro-oncogenic Phenotype in HNSCC. *Mol. Oncol.* 14 (12), 3065–3082. doi:10.1002/1878-0261.12806
- Gridelli, C., Rossi, A., Carbone, D. P., Guarize, J., Karachaliou, N., Mok, T., et al. (2015). Non-small-cell Lung Cancer. *Nat. Rev. Dis. Prim.* 1, 15009. doi:10.1038/nrdp.2015.9
- Hanahan, D., and Weinberg, R. A. (2011). Hallmarks of Cancer: the Next Generation. *Cell* 144 (5), 646–674. doi:10.1016/j.cell.2011.02.013
- Hanahan, D. (2022). Hallmarks of Cancer: New Dimensions. *Cancer Discov.* 12 (1), 31–46. doi:10.1158/2159-8290.CD-21-1059
- Herbst, R. S., Morgensztern, D., and Boshoff, C. (2018). The Biology and Management of Non-small Cell Lung Cancer. *Nature* 553 (7689), 446–454. doi:10.1038/nature25183
- Ingram, K., Samson, S. C., Zewdu, R., Zitnay, R. G., Snyder, E. L., and Mendoza, M. C. (2022). NKK2-1 Controls Lung Cancer Progression by Inducing DUSP6 to Dampen ERK Activity. *Oncogene* 41 (2), 293–300. doi:10.1038/s41388-021-02076-x
- Jianfeng, W., Yutao, W., and Jianbin, B. (2021). Long Non-Coding RNAs Correlate with Genomic Stability in Prostate Cancer: A Clinical Outcome and Survival Analysis. *Genomics* 113 (5), 3141–3151. doi:10.1016/j.ygeno.2021.06.029
- Jiang, H., Ma, B., Xu, W., Luo, Y., Wang, X., Wen, S., et al. (2021). A Novel Three-lncRNA Signature Predicts the Overall Survival of HNSCC Patients. *Ann. Surg. Oncol.* 28 (6), 3396–3406. doi:10.1245/s10434-020-09210-1
- Lee, A. J., Endesfelder, D., Rowan, A. J., Walther, A., Birkbak, N. J., Futreal, P. A., et al. (2011). Chromosomal Instability Confers Intrinsic Multidrug Resistance. *Cancer Res.* 71 (5), 1858–1870. doi:10.1158/0008-5472.CAN-10-3604
- Lee, S., Kopp, F., Chang, T. C., Sataluri, A., Chen, B., Sivakumar, S., et al. (2016). Noncoding RNA NORAD Regulates Genomic Stability by Sequestering PUMILIO Proteins. *Cell* 164 (1–2), 69–80. doi:10.1016/j.cell.2015.12.017
- Li, C., Liu, J., Lin, J., Li, Z., Shang, X., and Wang, H. (2019a). Poor Survival of Non-small-cell Lung Cancer Patients with Main Bronchus Tumor: a Large Population-Based Study. *Future Oncol.* 15 (24), 2819–2827. doi:10.2217/fon-2019-0098
- Li, R., Yang, Y. E., Yin, Y. H., Zhang, M. Y., Li, H., and Qu, Y. Q. (2019b). Methylation and Transcriptome Analysis Reveal Lung Adenocarcinoma-specific Diagnostic Biomarkers. *J. Transl. Med.* 17 (1), 324. doi:10.1186/s12967-019-2068-z
- Liang, Y., Ye, F., Cheng, Z., Ou, Y., Zou, L., Hu, Y., et al. (2021). Calculated Identification of Mutator-Derived lncRNA Signatures of Genomic Instability to Predict the Clinical Outcome of Muscle-Invasive Bladder Cancer. *Cancer Cell Int.* 21 (1), 476. doi:10.1186/s12935-021-02185-3
- Lim, S., Quinton, R. J., and Ganem, N. J. (2016). Nuclear Envelope Rupture Drives Genome Instability in Cancer. *Mol. Biol. Cell* 27 (21), 3210–3213. doi:10.1091/mbc.E16-02-0098
- Liu, P., Morrison, C., Wang, L., Xiong, D., Vedell, P., Cui, P., et al. (2012). Identification of Somatic Mutations in Non-small Cell Lung Carcinomas Using Whole-Exome Sequencing. *Carcinogenesis* 33 (7), 1270–1276. doi:10.1093/carcin/bgs148
- Lu, N., Liu, J., Xu, M., Liang, J., Wang, Y., Wu, Z., et al. (2021). CSMD3 Is Associated with Tumor Mutation Burden and Immune Infiltration in Ovarian Cancer Patients. *Int. J. Gen. Med.* 14, 7647–7657. doi:10.2147/IJGM.S335592
- Lukow, D. A., Sausville, E. L., Suri, P., Chunduri, N. K., Wieland, A., Leu, J., et al. (2021). Chromosomal Instability Accelerates the Evolution of Resistance to Anti-cancer Therapies. *Dev. Cell* 56, 2427–2439. doi:10.1016/j.devcel.2021.07.009
- Maimaiti, A., Wang, X., Pei, Y., Nuermaiti, N., Tuersunniyazi, A., Abula, Y., et al. (2021). Identification and Validation of a Novel Eight Mutant-Derived Long Non-coding RNAs Signature as a Prognostic Biomarker for Genome Instability in Low-Grade Glioma. *Aging (Albany NY)* 13 (11), 15164–15192. doi:10.18632/aging.203079
- Majidinia, M., and Yousefi, B. (2017). DNA Repair and Damage Pathways in Breast Cancer Development and Therapy. *DNA Repair (Amst)* 54, 22–29. doi:10.1016/j.dnarep.2017.03.009
- Malta, T. M., Sokolov, A., Gentles, A. J., Burzykowski, T., Poisson, L., Weinstein, J. N., et al. (2018). Machine Learning Identifies Stemness Features Associated with Oncogenic Dedifferentiation. *Cell* 173 (2), 338–354.e15. doi:10.1016/j.cell.2018.03.034
- Nair, L., Chung, H., and Basu, U. (2020). Regulation of Long Non-coding RNAs and Genome Dynamics by the RNA Surveillance Machinery. *Nat. Rev. Mol. Cell Biol.* 21 (3), 123–136. doi:10.1038/s41580-019-0209-0
- Negrini, S., Gorgoulis, V. G., and Halazonetis, T. D. (2010). Genomic Instability—Aan Evolving Hallmark of Cancer. *Nat. Rev. Mol. Cell Biol.* 11 (3), 220–228. doi:10.1038/nrm2858
- Nguyen, B., Fong, C., Luthra, A., Smith, S. A., DiNatale, R. G., Nandakumar, S., et al. (2022). Genomic Characterization of Metastatic Patterns from Prospective Clinical Sequencing of 25,000 Patients. *Cell* 185 (3), 563–575.e11. doi:10.1016/j.cell.2022.01.003
- Osrodek, M., and Wozniak, M. (2021). Targeting Genome Stability in Melanoma—A New Approach to an Old Field. *Int. J. Mol. Sci.* 22 (7), 3485. doi:10.3390/ijms22073485
- Peng, B., Li, H., Na, R., Lu, T., Li, Y., Zhao, J., et al. (2021). Identification of a Novel Prognostic Signature of Genome Instability-Related lncRNAs in Early Stage Lung Adenocarcinoma. *Front. Cell Dev. Biol.* 9, 706454. doi:10.3389/fcell.2021.706454
- Raynaud, F., Mina, M., Tavernari, D., and Ciriello, G. (2018). Pan-cancer Inference of Intra-tumor Heterogeneity Reveals Associations with Different Forms of Genomic Instability. *PLoS Genet.* 14 (9), e1007669. doi:10.1371/journal.pgen.1007669
- Reis, A. H., Vargas, F. R., and Lemos, B. (2016). Biomarkers of Genome Instability and Cancer Epigenetics. *Tumour Biol.* 37 (10), 13029–13038. doi:10.1007/s13277-016-5278-5
- Rummel, S., Valente, A. L., Kane, J. L., Shriver, C. D., and Ellsworth, R. E. (2012). Genomic (In)stability of the Breast Tumor Microenvironment. *Mol. Cancer Res.* 10 (12), 1526–1531. doi:10.1158/1541-7786.MCR-12-0425
- Sasaki, H., Hikosaka, Y., Okuda, K., Kawano, O., Moriyama, S., Yano, M., et al. (2010). NFE2L2 Gene Mutation in Male Japanese Squamous Cell Carcinoma of the Lung. *J. Thorac. Oncol.* 5 (6), 786–789. doi:10.1097/JTO.0b013e3181db3dd3
- Sen, M., Kindsfather, A., Danilova, L., Zhang, F., Colombo, R., LaPorte, M. G., et al. (2020). PTPRT Epigenetic Silencing Defines Lung Cancer with STAT3 Activation and Can Direct STAT3 Targeted Therapies. *Epigenetics* 15 (6–7), 604–617. doi:10.1080/15592294.2019.1676597
- Shao, J., Zhang, B., Kuai, L., and Li, Q. (2021). Integrated Analysis of Hypoxia-Associated lncRNA Signature to Predict Prognosis and Immune Microenvironment of Lung Adenocarcinoma Patients. *Bioengineered* 12 (1), 6186–6200. doi:10.1080/21655979.2021.1973874
- Sieber, O. M., Heinemann, K., and Tomlinson, I. P. (2003). Genomic Instability—The Engine of Tumorigenesis? *Nat. Rev. Cancer* 3 (9), 701–708. doi:10.1038/nrc1170
- Skoulidis, F., and Heymach, J. V. (2019). Co-occurring Genomic Alterations in Non-small-cell Lung Cancer Biology and Therapy. *Nat. Rev. Cancer* 19 (9), 495–509. doi:10.1038/s41568-019-0179-8

- Song, J. Y., Li, X. P., Qin, X. J., Zhang, J. D., Zhao, J. Y., and Wang, R. (2020). A Fourteen-lncRNA Risk Score System for Prognostic Prediction of Patients with Non-small Cell Lung Cancer. *Cancer Biomark.* 29 (4), 493–508. doi:10.3233/CBM-190505
- Sonugür, F. G., and Akbulut, H. (2019). The Role of Tumor Microenvironment in Genomic Instability of Malignant Tumors. *Front. Genet.* 10, 1063. doi:10.3389/fgene.2019.01063
- Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., et al. (2021). Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J. Clin.* 71 (3), 209–249. doi:10.3322/caac.21660
- Thompson, S. L., Bakhoum, S. F., and Compton, D. A. (2010). Mechanisms of Chromosomal Instability. *Curr. Biol.* 20 (6), R285–R295. doi:10.1016/j.cub.2010.01.034
- Wang, X., Wu, B., Yan, Z., Wang, G., Chen, S., Zeng, J., et al. (2021). Association of PTPRD/PTPRT Mutation with Better Clinical Outcomes in NSCLC Patients Treated with Immune Checkpoint Blockades. *Front. Oncol.* 11, 650122. doi:10.3389/fonc.2021.650122
- Wen, X. M., Xu, Z. J., Jin, Y., Xia, P. H., Ma, J. C., Qian, W., et al. (2021). Association Analyses of TP53 Mutation with Prognosis, Tumor Mutational Burden, and Immunological Features in Acute Myeloid Leukemia. *Front. Immunol.* 12, 717527. doi:10.3389/fimmu.2021.717527
- Whiteside, T. L. (2008). The Tumor Microenvironment and its Role in Promoting Tumor Growth. *Oncogene* 27 (45), 5904–5912. doi:10.1038/onc.2008.271
- Wu, G., Wang, Q., Zhu, T., Fu, L., Li, Z., Wu, Y., et al. (2021a). Identification and Validation of Immune-Related LncRNA Prognostic Signature for Lung Adenocarcinoma. *Front. Genet.* 12, 681277. doi:10.3389/fgene.2021.681277
- Wu, Y., Ni, H., Yang, D., Niu, Y., Chen, K., Xu, J., et al. (2021b). Driver and Novel Genes Correlated with Metastasis of Non-small Cell Lung Cancer: A Comprehensive Analysis. *Pathol. Res. Pract.* 224, 153551. doi:10.1016/j.prp.2021.153551
- Xu, M., Gong, S., Li, Y., Zhou, J., Du, J., Yang, C., et al. (2020). Identifying Long Non-coding RNA of Prostate Cancer Associated with Radioresponse by Comprehensive Bioinformatics Analysis. *Front. Oncol.* 10, 498. doi:10.3389/fonc.2020.00498
- Yan, K., Wang, Y., Shao, Y., and Xiao, T. (2021). Gene Instability-Related lncRNA Prognostic Model of Melanoma Patients via Machine Learning Strategy. *J. Oncol.* 2021, 5582920. doi:10.1155/2021/5582920
- Yuan, F., Cao, X., Zhang, Y.-H., Chen, L., Huang, T., Li, Z., et al. (2022). Identification of Novel Lung Cancer Driver Genes Connecting Different Omics Levels with a Heat Diffusion Algorithm. *Front. Cell Dev. Biol.* 10, 825272. doi:10.3389/fcell.2022.825272
- Zappa, C., and Mousa, S. A. (2016). Non-small Cell Lung Cancer: Current Treatment and Future Advances. *Transl. Lung Cancer Res.* 5 (3), 288–300. doi:10.21037/tlcr.2016.06.07
- Zhang, Y., He, Q., Hu, Z., Feng, Y., Fan, L., Tang, Z., et al. (2016). Long Noncoding RNA LINP1 Regulates Repair of DNA Double-Strand Breaks in Triple-Negative Breast Cancer. *Nat. Struct. Mol. Biol.* 23 (6), 522–530. doi:10.1038/nsmb.3211
- Zhang, M., Zhuang, G., Sun, X., Shen, Y., Wang, W., Li, Q., et al. (2017). TP53 Mutation-Mediated Genomic Instability Induces the Evolution of Chemoresistance and Recurrence in Epithelial Ovarian Cancer. *Diagn. Pathol.* 12 (1), 16. doi:10.1186/s13000-017-0605-8

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

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Zhang, Liu, Chen and Zhang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.