



## OPEN ACCESS

EDITED BY  
Xingdong Chen,  
Fudan University, China

REVIEWED BY  
Aimin Jiang,  
The First Affiliated Hospital of Xi'an  
Jiaotong University, China  
Giuseppe Jurman,  
Bruno Kessler Foundation (FBK), Italy

\*CORRESPONDENCE  
Qing Zhou,  
bayyzq@sina.com

SPECIALTY SECTION  
This article was submitted to  
Computational Genomics,  
a section of the journal  
Frontiers in Genetics

RECEIVED 18 August 2022  
ACCEPTED 26 September 2022  
PUBLISHED 10 October 2022

CITATION  
Hao L, Chen Q, Chen X and Zhou Q  
(2022), Integrated analysis of bulk and  
single-cell RNA-seq reveals the role of  
MYC signaling in lung adenocarcinoma.  
*Front. Genet.* 13:1021978.  
doi: 10.3389/fgene.2022.1021978

COPYRIGHT  
© 2022 Hao, Chen, Chen and Zhou. 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.

# Integrated analysis of bulk and single-cell RNA-seq reveals the role of MYC signaling in lung adenocarcinoma

Lu Hao<sup>1</sup>, Qiuyan Chen<sup>1</sup>, Xi Chen<sup>2</sup> and Qing Zhou<sup>2\*</sup>

<sup>1</sup>Science and Education Department, Shenzhen Baoan Shiyan People's Hospital, Shenzhen, China, <sup>2</sup>Central Laboratory, The People's Hospital of Baoan Shenzhen, The Second Affiliated Hospital of Shenzhen University, Shenzhen, China

MYC is one of the well-known oncogenes, and its important role in cancer still remains largely unknown. We obtained lung adenocarcinoma (LUAD) multi-omics data including genome, transcriptome, and single-cell sequencing data from multiple cohorts. We calculated the GSVA score of the MYC target v1 using the ssGSEA method, and obtained the genes highly correlated with this score by Spearman correlation analysis. Subsequent hierarchical clustering divided these genes into two gene sets highly associated with MYC signaling (S1 and S2). Unsupervised clustering based on these genes divided the LUAD samples into two distinct subgroups, namely, the MYC signaling inhibition group (C1) and activation group (C2). The MCP counter package in R was used to assess tumor immune cell infiltration abundance and ssGSEA was used to calculate gene set scores. The scRNA-seq was used to verify the association of MYC signaling to cell differentiation. We observed significant differences in prognosis, clinical characteristics, immune microenvironment, and genomic alterations between MYC signaling inhibition and MYC signaling activation groups. MYC-signaling is associated with genomic instability and can mediate the immunosuppressive microenvironment and promote cell proliferation, tumor stemness. Moreover, MYC-signaling activation is also subject to complex post-transcriptional regulation and is highly associated with cell differentiation. In conclusion, MYC signaling is closely related to the genomic instability, genetic alteration and regulation, the immune microenvironment landscape, cell differentiation, and disease survival in LUAD. The findings of this study provide a valuable reference to revealing the mechanism of cancer-promoting action of MYC in LUAD.

## KEYWORDS

MYC, lung adenocarcinoma, prognosis, tumor immunity, tumor stemness, cell proliferation, cell differentiation

## Introduction

Lung cancer is the most common malignant tumor of the respiratory system, and the basic and clinical research on lung cancer is increasingly attracting attention (Mok et al., 2019; Song et al., 2020; Song et al., 2022). Although the current research on the pathogenesis of lung cancer has made great progress, but the clinical treatment effect of lung cancer is still not satisfactory, and the long-term survival rate of lung cancer still has great room for improvement. With the deepening of research, molecular biology has been widely used in the field of lung cancer research, which not only provides many new methods for lung cancer research, but also makes the diagnosis and treatment of lung cancer into a new stage. According to the different biological characteristics, lung cancer is often divided into small cell lung cancer (SCLC) and non-small cell lung cancer (NSCLC) in clinical, among which the latter accounts for about 85% of all lung cancer patients (Cheng et al., 2019). Lung adenocarcinoma (LUAD) is the most common pathological subtype of lung cancer (Kim et al., 2019; Song et al., 2021). Lung cancer is a highly heterogeneous tumor, and lung cancer occurrence is a multi-gene, multi-factor joint regulation, multi-stage and multi-step process (Rajagopalan et al., 2018). A large number of molecular abnormalities and the mechanism of action remain to be explored.

The MYC gene family and its products are involved in the regulation of cell growth, differentiation, and programmed death, and play important roles in the formation of various tumors (Vo et al., 2016). Previous studies have shown that MYC can affect the cell cycle progression, and its amplification and overexpression can lead to c-Myc proto-oncogene activation, which subsequently promotes tumorigenesis and progression (King et al., 2016; Lee et al., 2016). It can also regulate the expression of VEGF, to control the angiogenesis (Thompson et al., 2017). MYC, acting as a transcription factor, can regulate the expression of a large number of genes in tumors. It can act as an amplifier that globally upregulates the expression of protein-coding genes within cancer cells. So with a slight MYC expression disorder, it is possible to promote cancer cell evolution (Jing et al., 2016; Wang et al., 2019; Poh et al., 2019). Recent advances in high-throughput sequencing technologies, such as whole-genome sequencing, have allowed us to analyze tumors in unprecedented depth, especially with the single-cell sequencing (scRNA-seq) technologies that have emerged in recent years (Jiang et al., 2022; Becht et al., 2018; Zhang et al., 2021). Among them, scRNA-seq is a new technology for high-throughput sequencing of mRNA at the single-cell level, studying the overall level of gene expression for individual cells. Given the non-negligible and important role of MYC in cancer cell growth, proliferation, and differentiation, this study innovatively used LUAD multi-omics data from multiple

cohorts to systematically investigate the relevance of transcriptional profile expression, genome instability, genetic alteration and regulation, immune microenvironment landscape, cell differentiation, and disease survival in Halkmark MYC target V1 gene sets by integrating bulk and single-cell RNA sequencing data. Figure 1 showed the workflow of this study. This study indicated significant differences in prognosis, clinical characteristics, immune microenvironment, and genomic alterations between MYC signaling inhibition and MYC signaling activation groups. MYC-signaling is associated with genomic instability and can mediate the immunosuppressive microenvironment and promote cell proliferation, tumor stemness. Moreover, MYC-signaling activation is also subject to complex post-transcriptional regulation and is highly associated with cancer cell differentiation. Take together, the findings of this study provide a valuable reference to revealing the mechanism of cancer-promoting action of MYC in LUAD.

## Materials and methods

### Data sources and sample collection

Expression profile data (tpm, counts, miRNA isoform) and clinical information for TCGA-LUAD ( $n = 516$ ) were downloaded from the GDC (<https://portal.gdc.cancer.gov/>). To avoid batch effects, the counts and tpm data that we used were directly derived from the STAR-counts workflow type, and were subsequently log<sub>2</sub>-transformed on the TPM data. Mutation data and copy number variation (CNV) data for the TCGA-LUAD dataset were also downloaded from the cBioPortal ([www.cbioportal.org/](http://www.cbioportal.org/)). Three independent LUAD cohorts were collected from the GEO database (<https://www.ncbi.nlm.nih.gov/geo/>) as external validations, respectively, GSE68465 ( $n = 443$ ) (Shedden et al., 2008), GSE72094 ( $n = 442$ ) (Schabath et al., 2016), and GSE31210 ( $n = 226$ ) (Okayama et al., 2012; Yamauchi et al., 2012). For processing the GEO data, we refer to the method of (Song et al. 2022). The LUAD GEO dataset included in this study was mainly considered based on the sample size. The above three LUAD cohorts have a substantial number of cases, which is an important basis for their inclusion in this study. The clinicopathological parameters of LUAD patients in the TCGA and GEO cohorts should be provided in the Supplementary Table S1.

### Definition of signature genes

Pathways for MSigDB database were acquired using the “msigdb” package in R. The enrichment score for the pathway “HALLMARK\_MYC\_TARGETS\_V1” was calculated using the ssGSEA algorithm of the “GSVA” package in R, and

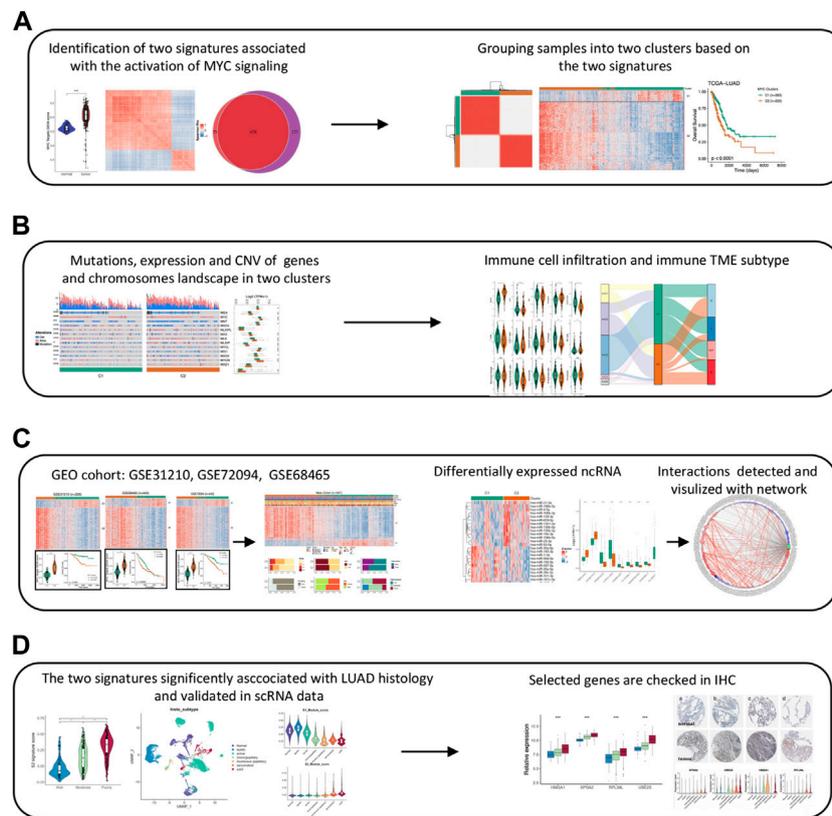


FIGURE 1

The work flow chart of this study. (A) Identifying MYC signaling related genes and clustering LUAD samples. (B) Analyzing the differences in multiple levels (Genome, transcriptome, and immune infiltration) between two clusters. (C) Validating the robustness of the two MYC signatures and constructing the network of differentially expressed lncRNAs and miRNAs. (D) Verifying signatures association with LUAD cell differentiation in scRNAseq data and IHC.

the genes highly correlated with this score were obtained by Spearman correlation analysis, with the threshold set as:  $Rho > 0.5$  and  $adj.p \text{ value} < 1e-3$ . Genes were subsequently filtered using univariate cox regression analysis and log rank test with  $p \text{ value} < 0.05$  as the threshold, and the two lists of genes obtained were set as intersection, and shared genes were considered as survival-related genes. We divided the resulting gene set into S1 signature set and S2 signature set by hierarchical clustering. All the genes in S1 were negatively correlated with HALLMARK\_MYC\_TARGETS\_V1 pathway score and  $HR < 1$ , while S2 was all positively correlated with HALLMARK\_MYC\_TARGETS\_V1 pathway score and  $HR > 1$ . Therefore, S1 signature was thought to be associated with MYC signaling inhibition, and S2 signature is associated with MYC signaling activation.

## Classifying samples with consensus clustering

Samples were consistently clustered using the “ConsensusClusterPlus” package in R (Qiu et al., 2021), with the parameters set to: distance = “euclidean”, clusterAlg = “km,” maxK = 5, reps = 100, pItem = 0.8, and the remaining parameters took the default values. And samples could be most distinctly classified when  $k = 2$ . After checking the expression level of the two signatures we previously identified in these two clusters, reasonably, we defined samples with highly expressed S1 signature genes as group C1 (MYC signaling inhibition group). Conversely, samples with highly expressed S2 signature genes were defined as group C2 (MYC signaling activation group).

## Analysis of the genomic variability

We used “data\_mutations\_extended.txt” downloaded from ciBioPortal to analyse the mutation landscape of two clusters. Non-silenced SNV was analyzed using the “maftools” R package. We focused on MYC gene family (MYC, MYCN, MYCL) and pathway core genes and genes listed as cancer driver genes by OncoKB (<https://www.oncokb.org/cancerGenes>). The Fisher test of genes mutated in at least 30 samples were also performed using the mafCompare algorithm to yield genes with significant differences in mutation frequency in the two groups. Copy number variations of related genes were analyzed using “data\_CNA.txt” data from ciBioPortal. Among them, the CNV state of genes is divided into -2, -1, 0, 1, 2, and 0 represents no CNV, 1 and 2 represent copy number amplification, and -1 and -2 represent copy number loss. Copy number variation at chromosome level were directly extracted from the data\_clinical\_sample.txt, and only the top 10 most significantly variated chromosome arms between C1 and C2 clusters were visualized. All the statistics of genomic variability were performed with two-sided Fisher’s exact test.

## Description of the tumor microenvironment

The scores of 10 typical immune cells, including T cells, CD8 T cells, CTL, B cells, NK, and monocytes, were calculated using MCP-counter. From a previous study (Bagaev et al., 2021), data including purity, intratumor heterogeneity, aneuploidy score, homologous recombination defects, BCR.Shannon, TCR.Shannon, M1/M2 macrophage were obtained. To further evaluate the impact of MYC on the immune microenvironment, we used TIDE (<http://tide.dfci.harvard.edu/>) to calculate the scores of TIL for MDSC, CAF, and M2, as well as two indicators related to immunotherapy response: T-cell dysfunction and exclusion (Jiang et al., 2018).

## Differential expression analysis of genes (including mRNA, lncRNA, miRNA) and the construction of CeRNA network

Using the “DESeq2” R package, the differential expression analysis was performed (Zhao et al., 2021). The threshold was set to adj.*p* value <0.001 and |log<sub>2</sub>FoldChange| > 0.5. And the resulting log<sub>2</sub>FC and adj.*p* value were used as the colors and sizes of the nodes in the subsequent network graph drawing, respectively. Circular nodes represents lncRNA, and square nodes represents miRNA. The selected lncRNA-miRNA

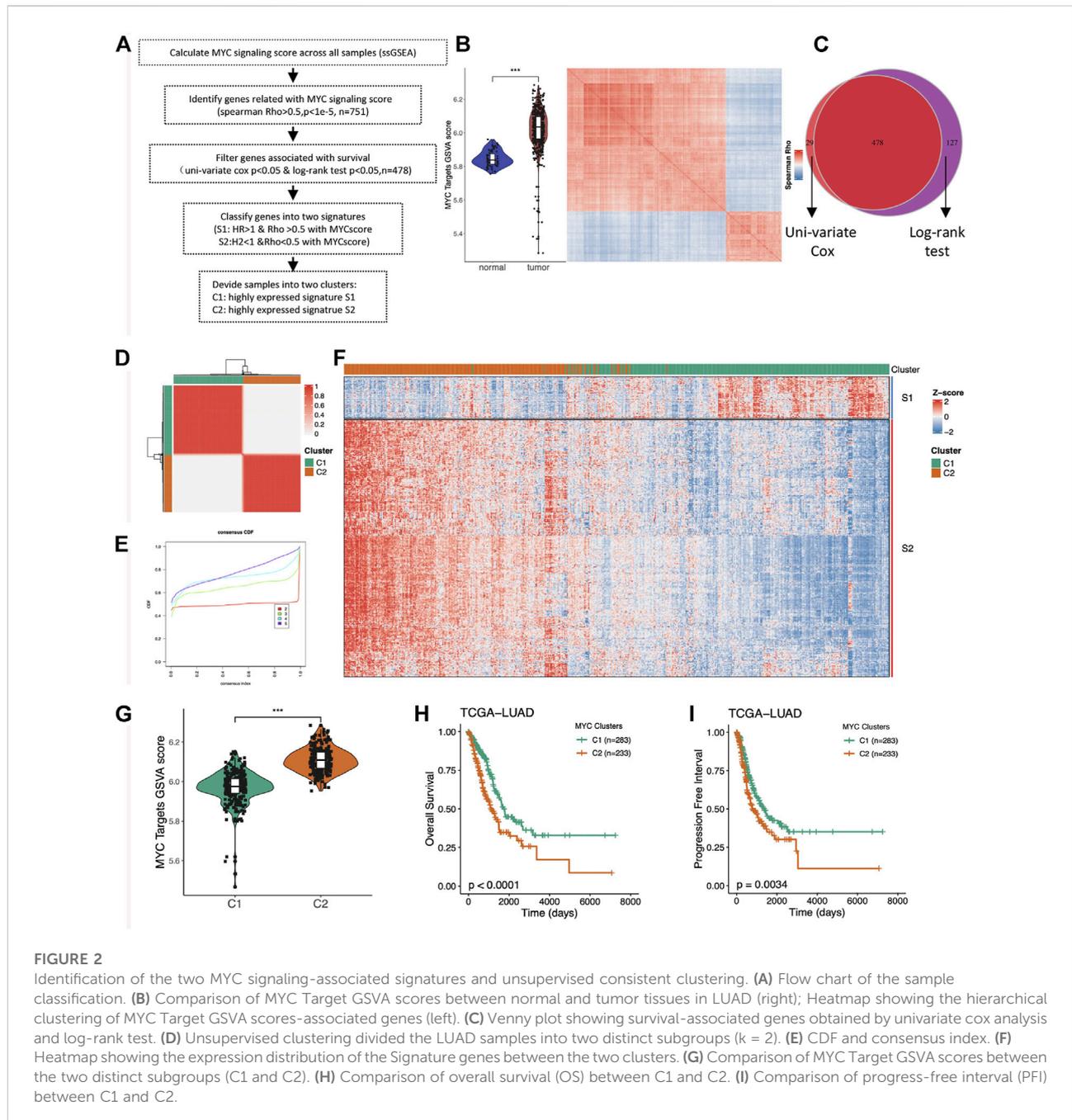
interaction, MYC/MYCN and-ncRNA interaction, and miRNA-MYC/MYCN interaction were predicted using the online tool RNAInter (<http://www.rnainter.org/>) and mirWalk (<http://mirwalk.umm.uni-heidelberg.de/>). For the prediction results, the drawing was performed using the “igraph” R package (Mora and Donaldson, 2011). Gray lines represent all possible interactions between ncRNA and MYC/MYCN, and red lines indicate possible interactions between ncRNA.

## Analysis of the scRNA-seq data

The expression matrix of scRNA-seq and the clinical information (such as histological type) of the samples were downloaded from the website (<https://doi.org/10.24433/CO.0121060.v1>) (Kumar and Song, 2022). The data contained a total of 114,489 cells from 10 LUAD samples and 10 normal lung tissue samples, and used 10x genomics for sequencing. Genes below expression in 100 cells were filtered out using the “Seurat” R package (Kumar and Song, 2022). Low-quality cells were filtered out by the criteria where the number of expressed genes was greater than 100 and less than 6,000 and the proportion of mitochondrial gene expression was less than 20. After defining and isolating epithelial cells from single cell expression profiling data of total cells, again, samples with less than 100 epithelial cells were filtered out for subsequent analysis. The top 15 principal components were used after PCA dimension reduction. Eventually we obtained 3,684 normal epithelial cells from the normal samples, and 15,477 malignant epithelial cells from the tumor samples. The signature module score was calculated using the AddModuleScore function.

## The interaction network of the Signature gene and immunohistochemical

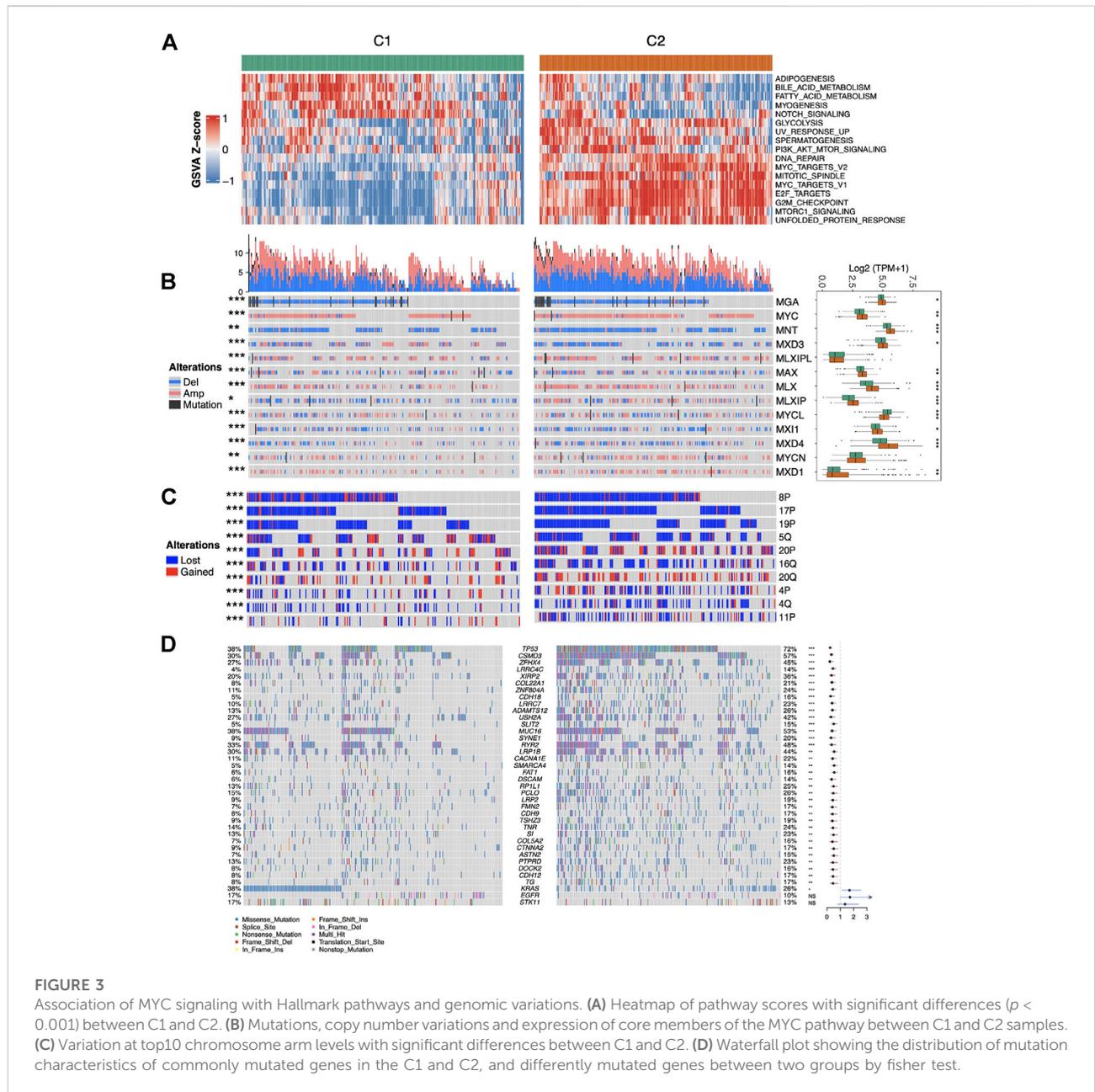
Genes with significant differential expression between the two groups (C1 and C2 groups) and the genes belonging to S1/S2 signature set were included in the potential nodes. Associations between nodes were obtained by correlation analysis, and edges with lower associations were filtered out. The visualization was then performed using the “igraph” R package. Circle size represents the -log<sub>10</sub> (*p*-value), and the circle color indicates the log<sub>2</sub>FoldChange for the difference analysis after C1/C2 grouping. Pictures of IHC staining derived from normal samples and LUAD samples were selected on the HPA website (<https://www.proteinatlas.org/>) to verify the relationship of key genes to cell differentiation. Here, “HPAanalyze” R package was used to download the high definition IHC pictures (Tran et al., 2019).



## Statistical analysis

All statistical analysis was done using R. Where the KM survival analysis was performed by log rank test using the “survival” and “survminer” R packages, and the univariate and multivariate cox were done using the basis function coxph. We

filtered out samples with less than 30 days of follow-up date before performing a survival analysis. Student’s *t*-Test was used to compare the differences in gene expression levels between clusters. A *p*-value of less than 0.05 was considered statistically significant. Heatmaps were all plotted using the “ComplexHeatmap” R package.



**FIGURE 3** Association of MYC signaling with Hallmark pathways and genomic variations. **(A)** Heatmap of pathway scores with significant differences ( $p < 0.001$ ) between C1 and C2. **(B)** Mutations, copy number variations and expression of core members of the MYC pathway between C1 and C2 samples. **(C)** Variation at top10 chromosome arm levels with significant differences between C1 and C2. **(D)** Waterfall plot showing the distribution of mutation characteristics of commonly mutated genes in the C1 and C2, and differently mutated genes between two groups by fisher test.

## Results

### Identification of the two MYC signaling-associated signatures

The level of MYC signaling activation cannot be simply judged by MYC gene expression and copy number amplification. Considering that the genes regulated by the same pathway are similar in their expression patterns, we assessed the degree of MYC signaling activation by looking at the overall expression levels of the MYC target genes, and

obtained all the highly correlated genes by similarity analysis. We performed the subsequent analysis as to **Figure 2A**. We first evaluated the enrichment score of the MYC\_TARGET\_V1 pathway by ssGSEA algorithm, then found the genes highly related with the score through correlation analysis, and performed hierarchical clustering (**Figure 2B**). As we expected, these genes could be divided into two groups that were highly concordant, with one group being highly positively correlated with MYC\_TARGET\_V1 and the other group being highly negatively correlated. To further screen for key genes, we further filtered out 478 survival-related genes by

intersection using univariate cox analysis and log rank test of the obtained genes (Figure 2C; Supplementary Table S2). Importantly, both cluster genes negatively associated with MYC\_TARGET\_V1 were associated with better prognosis, and both genes positively associated with MYC\_TARGET\_V1 were associated with worse prognosis. Therefore, we defined these two cluster gene sets as S1 and S2, respectively.

## The MYC signaling-associated signature could divide LUAD patients into two clinical clusters

Considering that the S1 and S2 genes have significantly different characteristics, we subsequently performed unsupervised consistent clustering of LUAD samples based on the expression of the signature genes, and finally obtained two clusters of samples (Figures 2D,E). One group of samples highly expressed the S1 signature gene, while the other group also highly expressed the S2 signature gene (Figure 2F), so we named it as the corresponding two C1 and C2 groups. Group C2 was the MYC signaling activation group, and group C1 was the MYC signaling inhibition group. The MYC scores were significantly different between the two groups (Figure 2G). In addition, we also found significant differences in OS (Figure 2H) and PFI (Figure 2I). This suggests important roles of MYC signaling in LUAD.

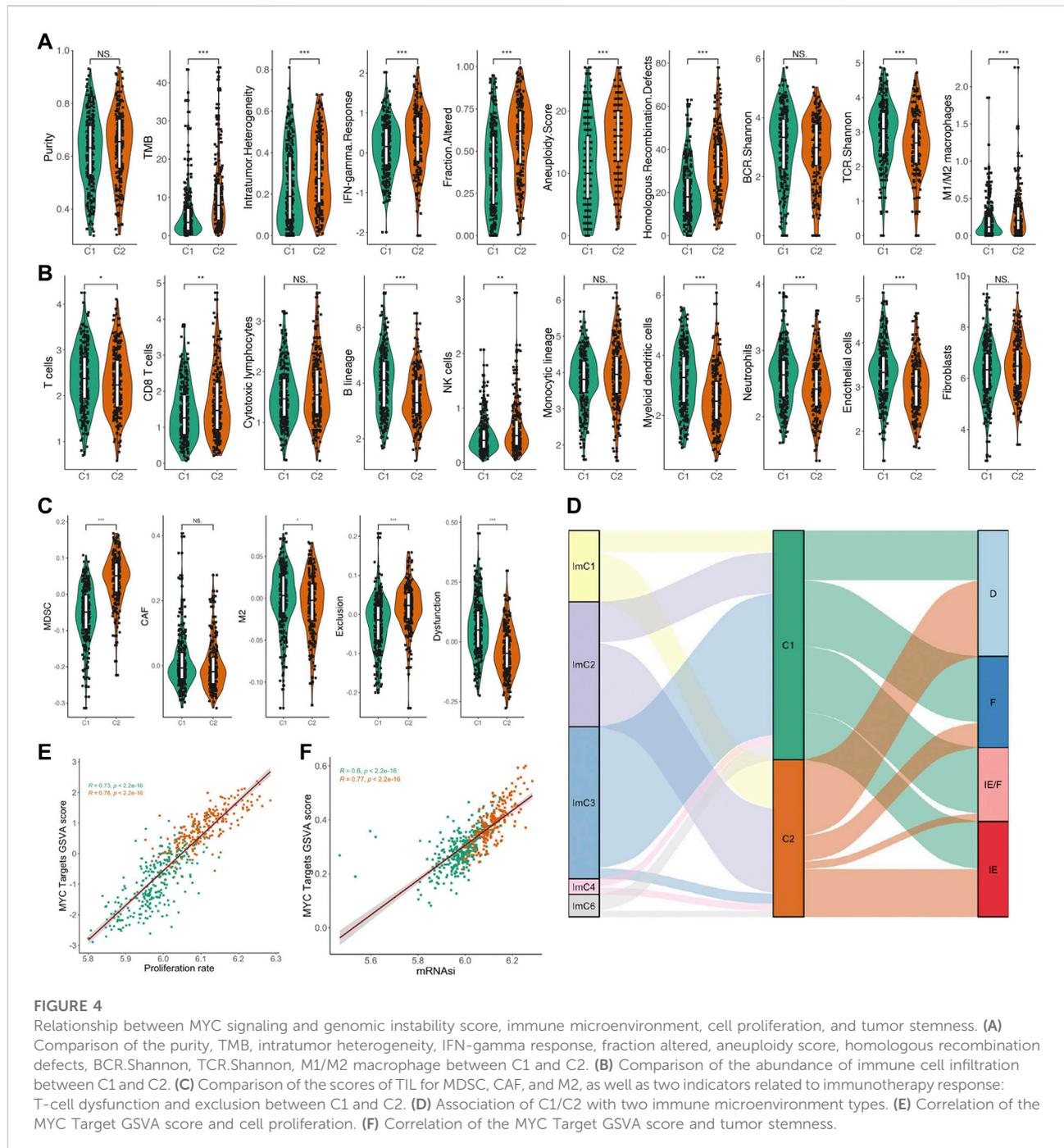
## Association of MYC signaling with Hallmark pathways and genomic variations

The association of MYC signaling with oncogenic pathways and genomic variants remains unclear, therefore, we investigated the GSEA score differences in Hallmark pathways between MYC signaling activation (C2) and inhibition (C1) groups. As shown in Figure 3A, in addition to the MYC and cell-cycle-related pathways, the pathways such as glycolysis and PI3K were also up-regulated in C2. Copy number variation (CNV) in all MYC pathway core genes were significantly different between C1 and C2 (fisher exact test  $p < 0.05$ ). Specifically, these genes developed CNV more frequently in C2, and MXD3 was both primarily lost in C2 and mostly amplified in C1. MLXIP was the opposite. This suggested that CNV changes were important causes of MYC pathway activation. Meanwhile, besides MLXIP and MYCN, other genes also differed in their expression between C1 and C2. Interestingly, although MXD3 experienced more copy number loss in C2, its expression remained higher in C2 (Figure 3B). We also examined CNV differences in chromosome levels between C1 and C2 (Figure 3C). Not surprisingly, multiple chromosomes-level CNV differences exist between C1 and C2. In addition to occurring more

frequently in C2, the types of variants occurring also varied, such as 5q being more amplified in C1. In LUAD, mutations in many key genes play a crucial role in tumor development. They are known as the driver genes. We examined the mutation situation between C1 and C2. The results showed that besides KRAS, EGFR, STK11 (these genes were thought to be mutually exclusive to MYC pathway activation in previous studies (Zhang et al., 2016; Mollaoglu et al., 2017)), most genes were more mutated and higher in C2 (Figure 3D). Overall, the results of this study indicate that MYC signaling is closely related with oncogenic pathways and genomic variants.

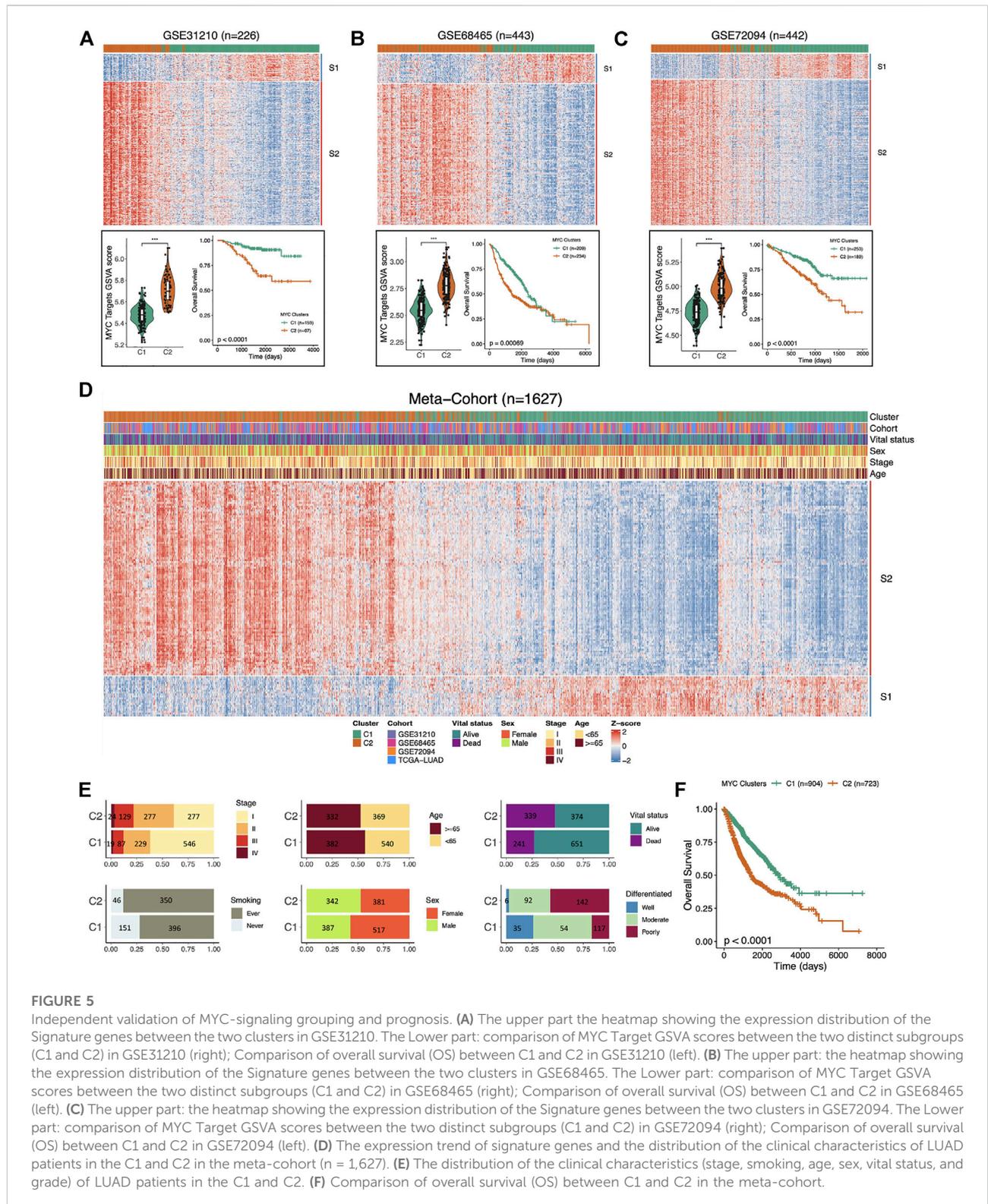
## MYC-signaling associates with genomic instability, mediates the immunosuppressive microenvironment, and promotes cell proliferation, and tumor stemness

In the above analysis, we found that the MYC signaling activation group was significantly different from the inhibition group in terms of genetic mutations. From this, we further investigated the differences in genomic instability scores between the two groups. We curated a list of genomic instability scores from a previous study (Bagaev et al., 2021), which was composed of the mutation burden score, the aneuploidy score, and the HRD score. The mutation burden score was non-silent mutations per Mb. The aneuploidy score reported the total number of arm-level amplifications and deletions and was computed using ABSOLUTE. Our results indicate that the MYC signaling activation group presents a higher genomic instability score than the MYC signaling inhibition group (Figure 4A). In addition, we also found that the MYC signaling activation group also showed higher intratumoral heterogeneity, IFN-gamma response and M1/M2 macrophages and lower TCR shannon, while the tumor purity and BCR shannon did not be significantly different between the two groups (Figure 4A). To analyze the effect of MYC signaling on the immune cells and the tumor microenvironment, we calculated the infiltration levels of the 10 immune cells using the MCP counter R package and performed a statistical test with a  $t$ -test (Figure 4B). We found that the cells mediating tumor killing (CD8<sup>+</sup> T cells, NK cells) had a higher infiltration abundance in the MYC signaling activation group. As important as the infiltration abundance of the immune cells in mediating the tumor immune response is the functional status of the immune cells, so we also evaluated indicators that reflect the immune function of LUAD with the online web tool TIDE (<http://tide.dfci.harvard.edu/>). The results showed that the tumor immune dysfunction score was significantly lower in the MYC signaling activation group when compared to the MYC signaling inhibition group (Figure 4C). This further suggests the importance of MYC signaling in mediating the tumor



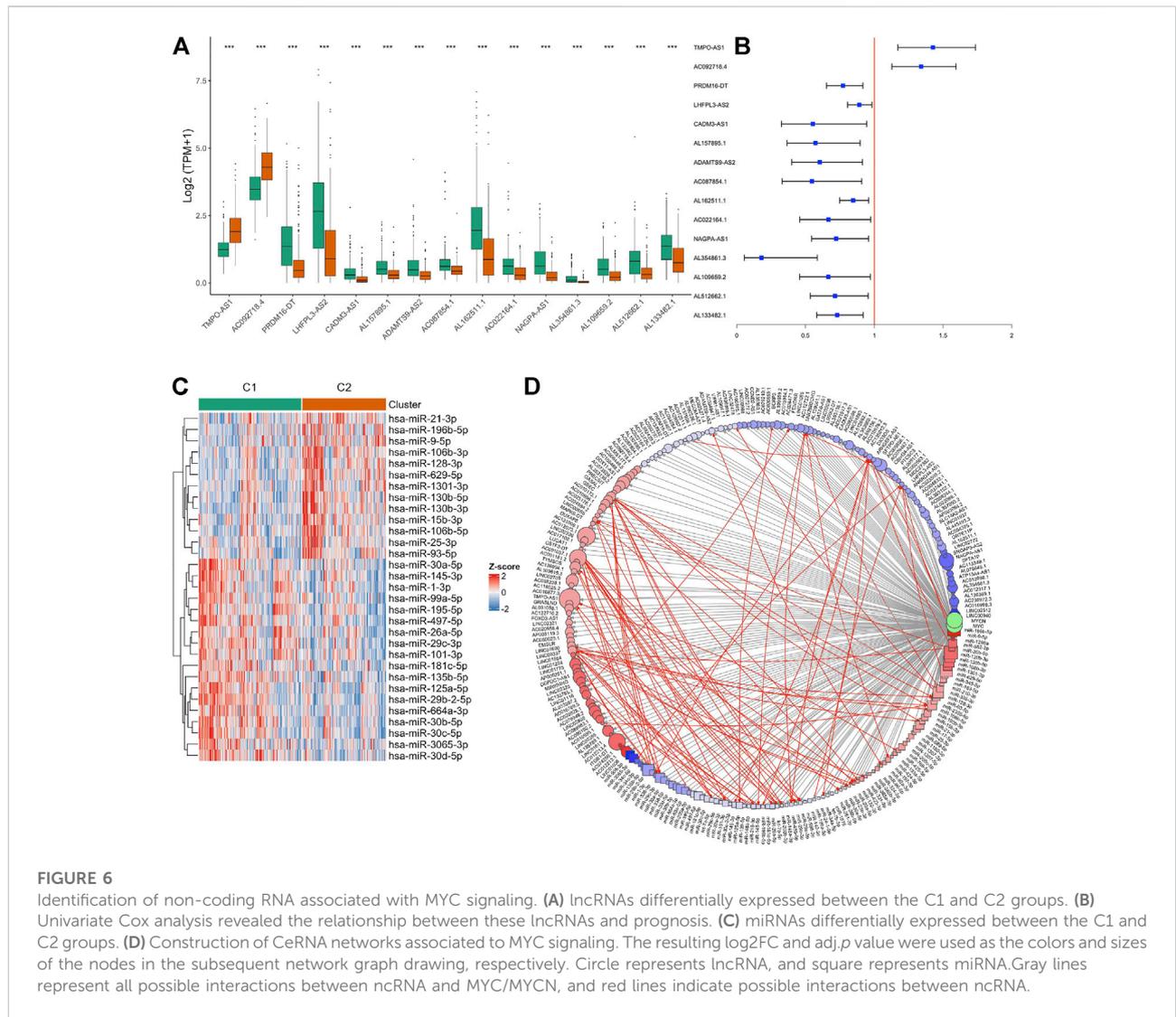
immunosuppressive microenvironment. For the tumor immune microenvironment, David Sacks et al. classified cancer samples into immune subtype in C1-C6 (Sacks et al., 2018). Similarly, Alexander Bagaev et al. defined the pan cancer sample of TCGA as four isoforms: IE, IE/F, D, and F (Bagaev et al., 2021). We explored the association between both C1/C2 groups and the tumor microenvironment of these two different differentiation

methods. Coincidentally, our data suggest that ImC3 has a largely overlapping relationship with C1 (Figure 4D). This further highlights the association of MYC signaling with the tumor immunosuppressive microenvironment. Incidentally, we also explored the relationship between MYC signaling and cell proliferation and tumor stemness. Surprisingly, the MYC score had a significant correlation with both (Figures 4E,F).



**FIGURE 5**

Independent validation of MYC-signaling grouping and prognosis. **(A)** The upper part the heatmap showing the expression distribution of the Signature genes between the two clusters in GSE31210. The Lower part: comparison of MYC Target GSVA scores between the two distinct subgroups (C1 and C2) in GSE31210 (right); Comparison of overall survival (OS) between C1 and C2 in GSE31210 (left). **(B)** The upper part: the heatmap showing the expression distribution of the Signature genes between the two clusters in GSE68465. The Lower part: comparison of MYC Target GSVA scores between the two distinct subgroups (C1 and C2) in GSE68465 (right); Comparison of overall survival (OS) between C1 and C2 in GSE68465 (left). **(C)** The upper part: the heatmap showing the expression distribution of the Signature genes between the two clusters in GSE72094. The Lower part: comparison of MYC Target GSVA scores between the two distinct subgroups (C1 and C2) in GSE72094 (right); Comparison of overall survival (OS) between C1 and C2 in GSE72094 (left). **(D)** The expression trend of signature genes and the distribution of the clinical characteristics of LUAD patients in the C1 and C2 in the meta-cohort (n = 1,627). **(E)** The distribution of the clinical characteristics (stage, smoking, age, sex, vital status, and grade) of LUAD patients in the C1 and C2. **(F)** Comparison of overall survival (OS) between C1 and C2 in the meta-cohort.

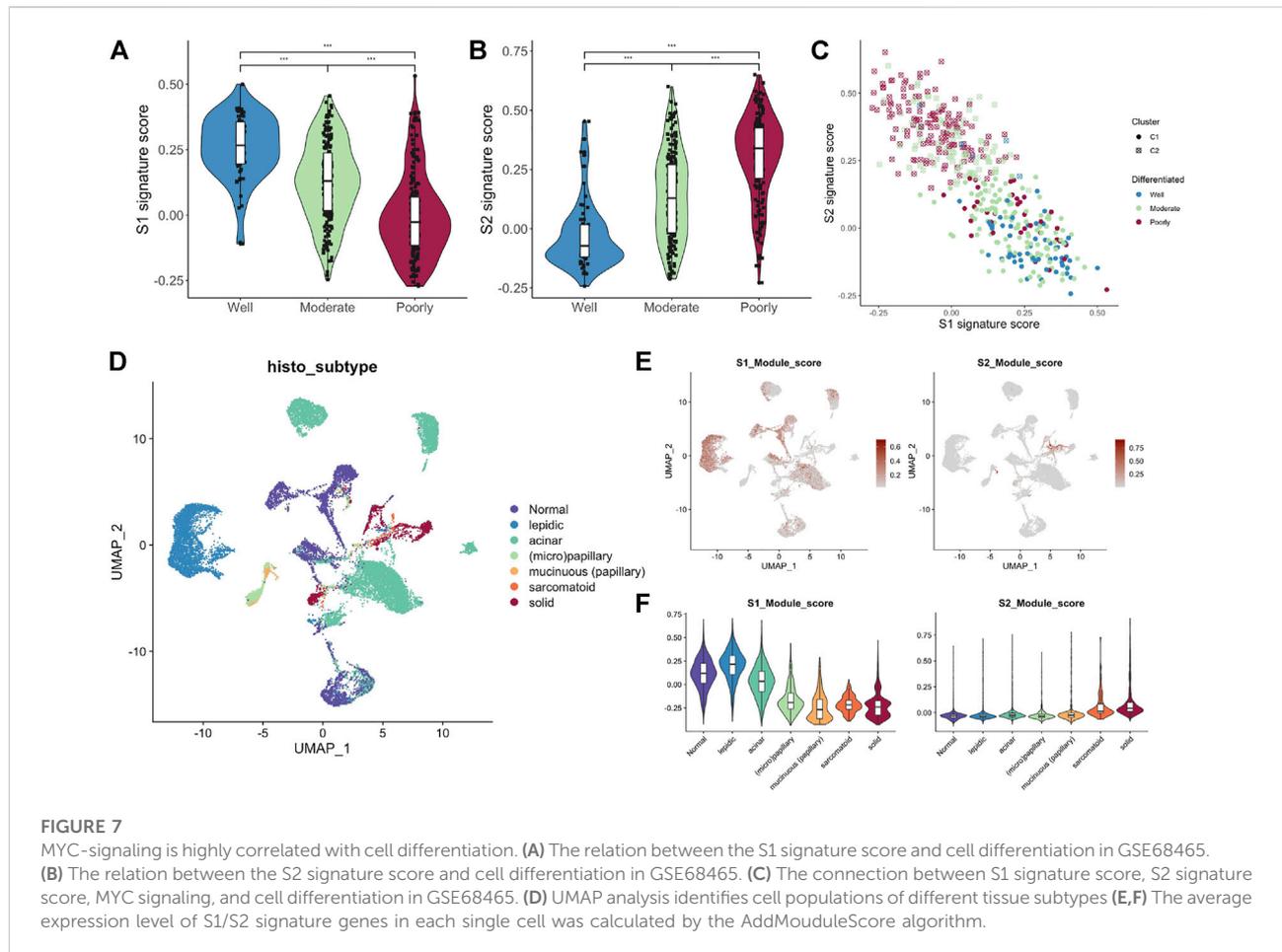


### Independent validation of MYC-signaling grouping and prognosis

To verify that the two signature (S1 and S2) we defined were stable on dividing LUAD samples into C1 and C2 groups according to MYC-signaling activation levels, we used three independent GEO datasets and a meta-cohort including 1,627 cases. The results showed that the MYC-signaling grouping was robust, which could efficiently classify samples into MYC-signaling activation group (C2) and MYC-signaling inhibition (C1), and were always highly correlated with patient prognosis (Figures 5A–D). Subsequently, we also investigated the distribution of clinical characteristics between the two groups, and we found that the MYC signaling activation group had more dead patients, who had later staging and poor cell differentiation, as shown in Figure 5E. Furthermore, Figure 5F also further confirmed that MYC C2 patients had a shorter OS.

### MYC-signaling activation was subject to complex post-transcriptional regulation

In both the TCGA and GSE31210 data, some samples were still classified into the MYC-signaling inhibition group (C1) even with MYC experiencing copy number amplification. This suggested that MYC-signaling activation was complex regulated. Coincidentally, we found 15 lncRNAs in these two signature gene sets (S1 and S2), of which 13 belong to S1 and 2 belong to S2 (Figure 6A). And the univariate cox analysis suggested that they were all associated with prognosis (Figure 6B). In addition to lncRNA, miRNA may also play an important role in regulating MYC-signaling activation. Therefore, we also explored the differentially expressed miRNAs between the two groups (Figure 6C), and found 31 miRNAs were differentially expressed between groups. lncRNA and miRNA, mRNA may regulate gene expression



through the CeRNA mechanism, and may also independently affect protein expression through other mechanisms such as acetylation. So we constructed a potential MYC/MYCN expression regulatory network (Figure 6D).

## MYC-signaling was highly correlated with cell differentiation

Histologically, samples with highly differentiated tumor cells were highly concentrated in C1, while those with poorly differentiated cells were highly concentrated in C2. Moreover, the S1/S2 signature score can independently distinguish the tumor cell differentiation level (Figures 7A–C). To further test the significance of these two signature, epithelial cells from LUAD samples and normal lung tissue at different differentiation levels were isolated and analyzed separately. After dimensionality reduction by PCA and UMAP, we obtained 3,684 normal epithelial cells, and 15,477 malignant epithelial cells (Figure 7D). The average expression level of S1/

S2 signature in each single cell was calculated by the AddMouduleScore algorithm (Figures 7E,F). The results were highly consistent with the previous findings. We found eight genes that were highly associated with cell differentiation were significantly differentially expressed in samples with different levels of differentiation in GSE68465 and showed consistent changes with the degree of differentiation (Figure 8A). Among them, CYP4B1, SUSD2, NFIX, and SYNE1 were highly expressed in normal lung epithelial cells and highly differentiated epithelial cells (Figures 8B–E bottom). The IHC staining also indicated that they had a higher expression in the normal (Figures 8B–E top) relative to the LUAD samples (Figures 8B–E middle). KPNA2, UBE2S, HMGA1, and RPL39L were highly expressed in poorly differentiated lung epithelial cells (Figures 9A–D bottom). The IHC staining also indicated that they had a lower expression in the normal (Figures 9A–D top) relative to the LUAD samples (Figures 9A–D middle). These results indicated that MYC-signaling was highly correlated with cancer cell differentiation.

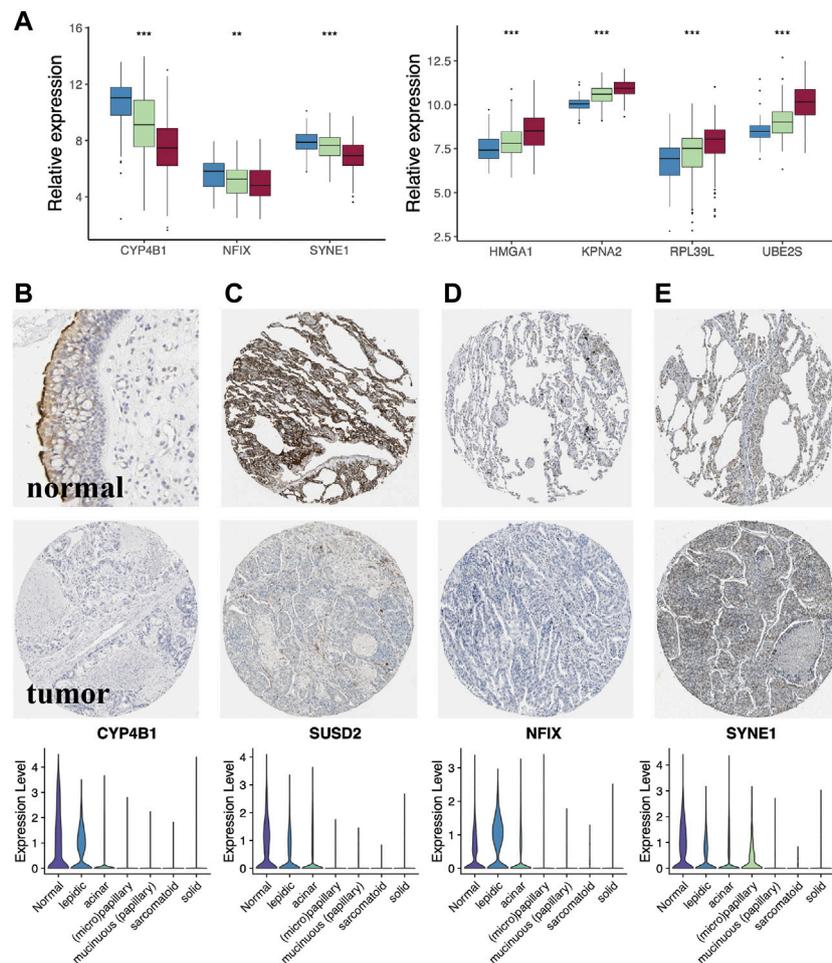


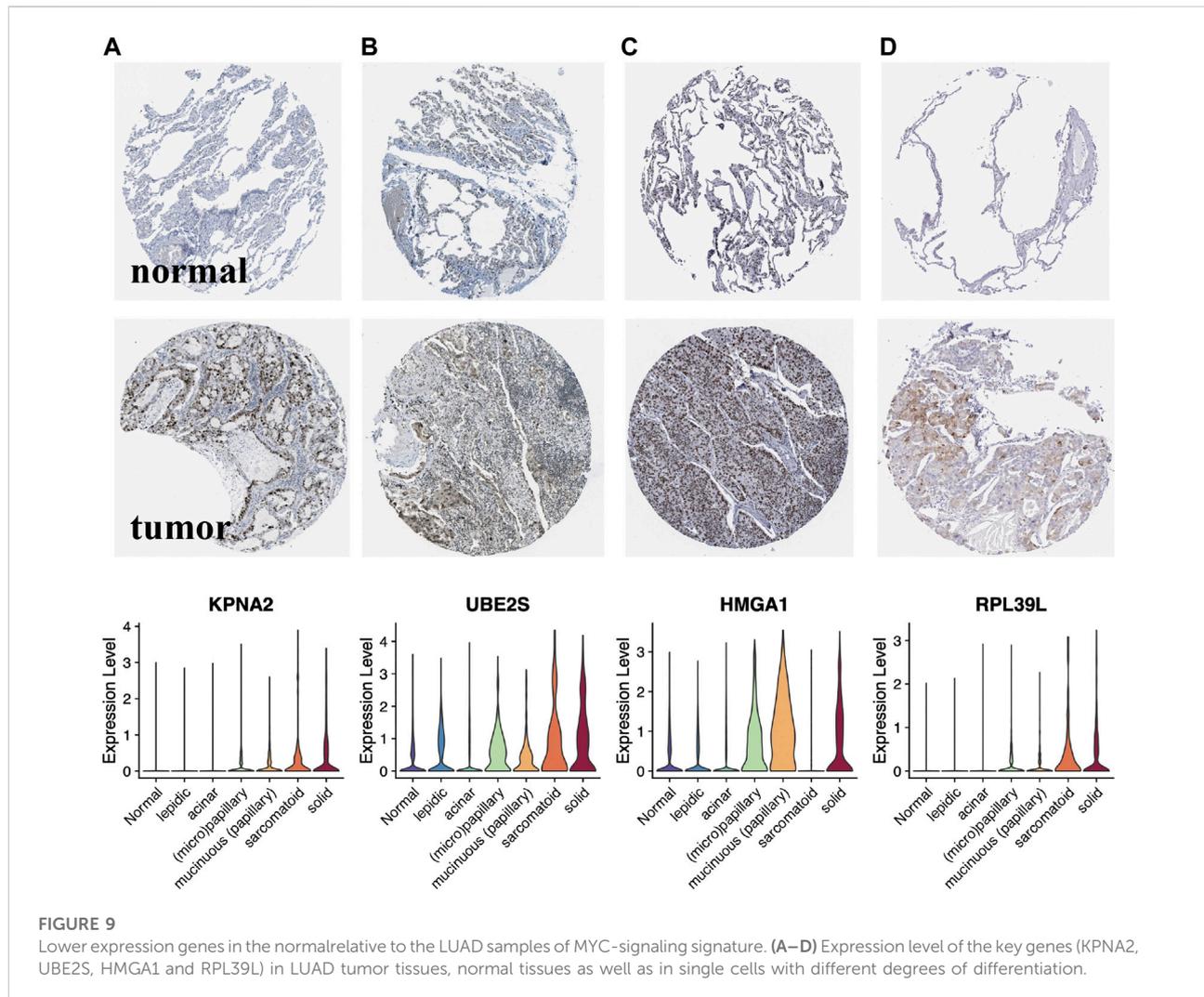
FIGURE 8

The key genes of MYC-signaling signature. (A) Genes highly associated with tumor cell differentiation. (B–E) Expression level of the key genes (CYP4B1, SUSD2, NFIX and SYNE1) in LUAD tumor tissues, normal tissues as well as in single cells with different degrees of differentiation.

## Discussion

In this study, we examined the tumor MYC\_TARGET\_V1 score in multiple large LUAD cohorts, and its correlation with transcriptional profile expression, genomic instability, genetic alteration and regulation, immune microenvironment landscape, cell differentiation, and disease survival. MYC, acting as a transcription factor, and a slight disturbance of MYC expression may promote cancer cell evolution. To investigate the level of MYC signaling activation, we analyzed the expression levels of the MYC gene family and pathway core genes. We found that these genes were mostly significantly different between the two groups. Further investigating the copy number variation of the core MYC pathway genes between the two groups, we found that they did not show significant differences in the copy number variation. This implies that the activation of MYC signaling is

epigenetically regulated, for example, DNA methylation (Panopoulos et al., 2017). It has been shown that the turnover of Myc proteins is determined by a cascade of phosphorylation and ubiquitination events (Liu et al., 2019; Parang et al., 2017). Notably, there is still a lack of evidence on whether MYC is regulated by ncRNA. In contrast, MYC, as a transcription factor, can regulate the activation and expression of ncRNA, for example, the miR-15 and let-7 (Adams and Eischen, 2016). In the study from Hou et al. (Zhang et al., 2022), they found that the MYC/MAX-trans-activated LINC00958 could promote the malignant behavior of LUAD by recruiting HOXA1 and inducing oncogenic reprogramming. To further clarify the pathways in which MYC is involved, we calculated the enrichment scores of the 50 Hallmark pathways in MsigDB by the ssGSEA algorithm, and found differences in multiple Hallmark pathway enrichment scores between the two groups, a finding that was also



consistent with previous studies. *c-Myc*, an important member of the MYC gene family, acts as a proto-oncogene localized to chromosome 8q24.1 and can be activated by chromosomal amplification, translocation, and rearrangement (Xu-Monette et al., 2016). In the above analysis, we also found that the MYC signaling activation group was significantly different from the inhibition group in terms of gene mutations. We further investigated the relationship between MYC signaling and the genome instability. We found that the MYC signaling activation group presented higher genomic instability scores. This result was not surprising, as reported in previous studies (Hao et al., 2016; H. Song et al., 2020).

More and more researchers have noticed the close link between tumor immune microenvironment and cancer occurrence and progression (Wang et al., 2016; Yuan et al., 2018; Tekpli et al., 2019). The MYC gene was also reported to be involved in the immune regulation of multiple tumors (Han et al., 2019; Swaminathan et al., 2020). In this study, we found

that the cells mediating the tumor-killing effect had a higher infiltration abundance in the MYC signaling activation group. Further investigating the functional status of the immune cells, we found that the tumor immune dysfunction score was significantly lower in the MYC signaling activation group as compared to the MYC signaling inhibition group. This further suggests the importance of MYC signaling in mediating the tumor immunosuppressive microenvironment. Moreover, the association of MYC with immune checkpoints is also slowly being revealed. For example, Thongsuksai et al. (Sunpaweravong et al., 2022) found that NSCLC tissues significantly express more *c-Myc* and PD-L1 compared to the matched normal respiratory epithelium, highlighting the important role of these key drivers in tumorigenesis. Laura Soucek and his colleagues (Masso-Valles et al., 2020) suggested that MYC, MYCL and MYCN might be therapeutic targets for lung cancer and that elevated Myc levels were also associated with treatment resistance, there may be significant opportunities for the combination of Myc inhibitors

with immunotherapies. It is well known that cancer occurrence is closely associated with the uncontrolled clonal proliferation of cells (Chung et al., 2019). As a well-known protooncogenic gene, MYC has been reported in mediating cell proliferation (Feist et al., 2018). However, its relationship between it and cell proliferation and tumor stemness in LUAD also needs to be further clarified. Our study showed a significant positive correlation between cell proliferation rate as well as tumor stemness and MYC score, and further highlights its non-negligible role in regulating LUAD cell proliferation and maintaining tumor stemness. Previous studies (Ireland et al., 2020; Patel et al., 2021) have revealed the key role of MYC in small cell lung cancer (SCLC) from a genomics perspective. Trudy G. Oliver et al. (Ireland et al., 2020) defined different SCLC molecular isoforms, based on the expression of ASCL1, NEUROD1, POU2F3, or YAP1. They used mouse and human models with time-series single-cell transcriptomic analysis to reveal the dynamic evolution of MYC-driven SCLC isoforms, finding that in neuroendocrine cells, MYC activated Notch to dedifferentiate tumor cells, promoting the temporal transition of SCLC from ASCL1 + to NEUROD1 + to YAP1 + state. The study by Hideo Watanabe and his colleagues (Patel et al., 2021) has also revealed the previously undescribed roles of the historically defined general oncogenes c-Myc and L-Myc for regulating lineage plasticity across molecular subtypes and histological subclasses. From the data currently available, MYC in SCLC seems to be studied more fully compared with LUAD. Therefore, it is still important to further investigate the potential role of MYC in LUAD from multi-omics data.

Overall, we used information from up to 1,600 samples of multiple LUAD cohorts to represent the important role of MYC signaling in LUAD from multiple dimensions of transcriptional profile expression, genomic instability, genetic alteration and regulation, immune microenvironment landscape, cell differentiation, and disease survival. This provides a valuable reference for deeply revealing the mechanism of cancer-promoting action of MYC in LUAD. However, like many other studies, the present study has some limitations. First, this study was a retrospective study and it was difficult to completely eliminate selective bias; second, although the important role of MYC in LUAD was described from multiple perspectives using multiple large study cohorts of LUAD and multiple bioinformatics approaches, further validation of the underlying experiments was lacking.

## 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

LH and QZ conceived and designed the study. LH and QZ analyzed the data. LH and QZ wrote and edited the manuscript. XC and QC participated in the study design and provided critical insights. All authors have approved the manuscript.

## Funding

This work was supported by Shenzhen Baoan Shiyuan People's Hospital Fund (2020SY07) and Science and Technology Bureau of Baoan (2021JD067).

## Acknowledgments

The authors hereby express their gratitude to all participants who supported the study, especially the database providers who provided the data for the analysis.

## 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.

## Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2022.1021978/full#supplementary-material>

## References

- Adams, C. M., and Eischen, C. M. (2016). Histone deacetylase inhibition reveals a tumor-suppressive function of MYC-regulated miRNA in breast and lung carcinoma. *Cell Death Differ.* 23 (8), 1312–1321. doi:10.1038/cdd.2016.9
- Bagaev, A., Kotlov, N., Nomie, K., Svekolkina, V., Gafurov, A., Isaeva, O., et al. (2021). Conserved pan-cancer microenvironment subtypes predict response to immunotherapy. *Cancer Cell* 39 (6), 845–865.e7. doi:10.1016/j.ccell.2021.04.014
- Becht, E., McInnes, L., Healy, J., Dutertre, C. A., Kwok, I., Ng, L. G., et al. (2018). Dimensionality reduction for visualizing single-cell data using UMAP. *Nat. Biotechnol.* 37, 38–44. doi:10.1038/nbt.4314
- Cheng, G., Zhang, Q., Pan, J., Lee, Y., Ouari, O., Hardy, M., et al. (2019). Targeting lisdiamine to mitochondria mitigates lung tumorigenesis and brain metastasis. *Nat. Commun.* 10 (1), 2205. doi:10.1038/s41467-019-10042-1
- Chung, S., Vail, P., Witkiewicz, A. K., and Knudsen, E. S. (2019). Coordinately targeting Cell-Cycle checkpoint functions in integrated models of pancreatic cancer. *Clin. Cancer Res.* 25 (7), 2290–2304. doi:10.1158/1078-0432.CCR-18-1620
- Feist, M., Schwarzfischer, P., Heinrich, P., Sun, X., Kemper, J., von Bonin, F., et al. (2018). Cooperative STAT/NF- $\kappa$ B signaling regulates lymphoma metabolic reprogramming and aberrant GOT2 expression. *Nat. Commun.* 9 (1), 1514. doi:10.1038/s41467-018-03803-x
- Han, H., Jain, A. D., Truica, M. I., Izquierdo-Ferrer, J., Anker, J. F., Lysy, B., et al. (2019). Small-molecule MYC inhibitors suppress tumor growth and enhance immunotherapy. *Cancer Cell* 36 (5), 483–497. doi:10.1016/j.ccell.2019.10.001
- Hao, T., Gaerig, V. C., and Brooks, T. A. (2016). Nucleic acid clamp-mediated recognition and stabilization of the physiologically relevant MYC promoter G-quadruplex. *Nucleic Acids Res.* 44 (22), 11013–11023. doi:10.1093/nar/gkw1006
- Ireland, A. S., Micinski, A. M., Kastner, D. W., Guo, B., Wait, S. J., Spainhower, K. B., et al. (2020). MYC drives temporal evolution of small cell lung cancer subtypes by reprogramming neuroendocrine fate. *Cancer Cell* 38 (1), 60–78. doi:10.1016/j.ccell.2020.05.001
- Jiang, A., Wang, J., Liu, N., Zheng, X., Li, Y., Ma, Y., et al. (2022). Integration of Single-Cell RNA sequencing and bulk RNA sequencing data to establish and validate a prognostic model for patients with lung adenocarcinoma. *Front. Genet.* 13, 833797. doi:10.3389/fgene.2022.833797
- Jiang, P., Gu, S., Pan, D., Fu, J., Sahu, A., Hu, X., et al. (2018). Signatures of T cell dysfunction and exclusion predict cancer immunotherapy response. *Nat. Med.* 24 (10), 1550–1558. doi:10.1038/s41591-018-0136-1
- Jing, H., Hu, J., He, B., Negron, A. Y., Stupinski, J., Weiser, K., et al. (2016). A SIRT2-Selective inhibitor promotes c-Myc oncoprotein degradation and exhibits broad anticancer activity. *Cancer Cell* 29 (4), 607. doi:10.1016/j.ccell.2016.03.011
- Kim, J. W., Marquez, C. P., Kostyrko, K., Koehne, A. L., Marini, K., Simpson, D. R., et al. (2019). Antitumor activity of an engineered decoy receptor targeting CLCF1-CNTFR signaling in lung adenocarcinoma. *Nat. Med.* 25 (11), 1783–1795. doi:10.1038/s41591-019-0612-2
- King, B., Boccalle, F., Moran-Crusio, K., Wolf, E., Wang, J., Kayembe, C., et al. (2016). The ubiquitin ligase Huwe1 regulates the maintenance and lymphoid commitment of hematopoietic stem cells. *Nat. Immunol.* 17 (11), 1312–1321. doi:10.1038/ni.3559
- Kumar, S., and Song, M. (2022). Overcoming biases in causal inference of molecular interactions. *Bioinformatics* 38, 2818–2825. doi:10.1093/bioinformatics/btac206
- Lee, J. K., Phillips, J. W., Smith, B. A., Park, J. W., Stoyanova, T., McCaffrey, E. F., et al. (2016). N-Myc drives neuroendocrine prostate cancer initiated from human prostate epithelial cells. *Cancer Cell* 29 (4), 536–547. doi:10.1016/j.ccell.2016.03.001
- Liu, P. Y., Tee, A. E., Milazzo, G., Hannan, K. M., Maag, J., Mondal, S., et al. (2019). The long noncoding RNA lncNBI promotes tumorigenesis by interacting with ribosomal protein RPL35. *Nat. Commun.* 10 (1), 5026. doi:10.1038/s41467-019-12971-3
- Masso-Valles, D., Beaulieu, M. E., and Soucek, L. (2020). MYC, MYCL, and MYCN as therapeutic targets in lung cancer. *Expert Opin. Ther. Targets* 24 (2), 101–114. doi:10.1080/14728222.2020.1723548
- Mok, T., Wu, Y. L., Kudaba, I., Kowalski, D. M., Cho, B. C., Turna, H. Z., et al. (2019). Pembrolizumab versus chemotherapy for previously untreated, PD-L1-expressing, locally advanced or metastatic non-small-cell lung cancer (KEYNOTE-042): A randomised, open-label, controlled, phase 3 trial. *Lancet* 393 (10183), 1819–1830. doi:10.1016/S0140-6736(18)32409-7
- Mollaoglu, G., Guthrie, M. R., Bohm, S., Bragelmann, J., Can, I., Ballieu, P. M., et al. (2017). MYC drives progression of small cell lung cancer to a variant neuroendocrine subtype with vulnerability to aurora kinase inhibition. *Cancer Cell* 31 (2), 270–285. doi:10.1016/j.ccell.2016.12.005
- Mora, A., and Donaldson, I. M. (2011). iRefR: An R package to manipulate the iRefIndex consolidated protein interaction database. *BMC Bioinforma.* 12, 455. doi:10.1186/1471-2105-12-455
- Okayama, H., Kohno, T., Ishii, Y., Shimada, Y., Shiraiishi, K., Iwakawa, R., et al. (2012). Identification of genes upregulated in ALK-positive and EGFR/KRAS/ALK-negative lung adenocarcinomas. *Cancer Res.* 72 (1), 100–111. doi:10.1158/0008-5472.CAN-11-1403
- Panopoulos, A. D., Smith, E. N., Arias, A. D., Shepard, P. J., Hishida, Y., Modesto, V., et al. (2017). Aberrant DNA methylation in human iPSCs associates with MYC-Binding motifs in a Clone-Specific manner independent of genetics. *Cell Stem Cell* 20 (4), 505–517. doi:10.1016/j.stem.2017.03.010
- Parang, B., Kaz, A. M., Barrett, C. W., Short, S. P., Ning, W., Keating, C. E., et al. (2017). BVES regulates c-Myc stability via PP2A and suppresses colitis-induced tumorigenesis. *Gut* 66 (5), 852–862. doi:10.1136/gutjnl-2015-310255
- Patel, A. S., Yoo, S., Kong, R., Sato, T., Sinha, A., Karam, S., et al. (2021). Prototypical oncogene family Myc defines unappreciated distinct lineage states of small cell lung cancer. *Sci. Adv.* 7 (5), eabc2578. doi:10.1126/sciadv.abc2578
- Poh, B., Koso, H., Momota, H., Komori, T., Suzuki, Y., Yoshida, N., et al. (2019). Foxr2 promotes formation of CNS-embryonal tumors in a Trp53-deficient background. *Neuro. Oncol.* 21 (8), 993–1004. doi:10.1093/neuonc/noz067
- Qiu, C., Shi, W., Wu, H., Zou, S., Li, J., Wang, D., et al. (2021). Identification of molecular subtypes and a prognostic signature based on Inflammation-Related genes in colon adenocarcinoma. *Front. Immunol.* 12, 769685. doi:10.3389/fimmu.2021.769685
- Rajagopalan, D., Tirado-Magallanes, R., Bhatia, S. S., Teo, W. S., Sian, S., Hora, S., et al. (2018). TIP60 represses activation of endogenous retroviral elements. *Nucleic Acids Res.* 46 (18), 9456–9470. doi:10.1093/nar/gky659
- Sacks, D., Baxter, B., Campbell, B., Carpenter, J. S., Cognard, C., Dippel, D., et al. (2018). Multisociety consensus quality improvement revised consensus statement for endovascular therapy of acute ischemic stroke. *Int. J. Stroke* 13 (6), 612–632. doi:10.1177/1747493018778713
- Schabath, M. B., Welsh, E. A., Fulp, W. J., Chen, L., Teer, J. K., Thompson, Z. J., et al. (2016). Differential association of STK11 and TP53 with KRAS mutation-associated gene expression, proliferation and immune surveillance in lung adenocarcinoma. *Oncogene* 35 (24), 3209–3216. doi:10.1038/onc.2015.375
- Shedden, K., Taylor, J. M., Enkemann, S. A., Tsao, M. S., Yeatman, T. J., Gerald, W. L., et al. (2008). Gene expression-based survival prediction in lung adenocarcinoma: A multi-site, blinded validation study. *Nat. Med.* 14 (8), 822–827. doi:10.1038/nm.1790
- Song, C., Guo, Z., Yu, D., Wang, Y., Wang, Q., Dong, Z., et al. (2020). A prognostic nomogram combining Immune-Related gene signature and clinical factors predicts survival in patients with lung adenocarcinoma. *Front. Oncol.* 10, 1300. doi:10.3389/fonc.2020.01300
- Song, C., Lu, Z., Lai, K., Li, D., Hao, B., Xu, C., et al. (2022). Identification of an inflammatory response signature associated with prognostic stratification and drug sensitivity in lung adenocarcinoma. *Sci. Rep.* 12 (1), 10110. doi:10.1038/s41598-022-14323-6
- Song, C., Wu, Z., Wang, Q., Wang, Y., Guo, Z., Li, S., et al. (2021). A combined Two-mRNA signature associated with PD-L1 and tumor mutational burden for prognosis of lung adenocarcinoma. *Front. Cell Dev. Biol.* 9, 634697. doi:10.3389/fcell.2021.634697
- Song, H., Liu, D., Dong, S., Zeng, L., Wu, Z., Zhao, P., et al. (2020). Epitranscriptomics and epiproteomics in cancer drug resistance: Therapeutic implications. *Signal Transduct. Target. Ther.* 5 (1), 193. doi:10.1038/s41392-020-00300-w
- Sunpaweravong, P., Thongwatchara, P., Chotipanvithayakul, R., Sangkhathat, S., and Thongsuksai, P. (2022). Expression and prognostic significance of c-Myc, ALK, ROS1, BRAF, and PD-L1 among patients with Non-Small cell lung cancer. *Clin. Med. Insights. Oncol.* 16, 11795549221092747. doi:10.1177/11795549221092747
- Swaminathan, S., Hansen, A. S., Heftdal, L. D., Dhanasekaran, R., Deutzmann, A., Fernandez, W., et al. (2020). MYC functions as a switch for natural killer cell-mediated immune surveillance of lymphoid malignancies. *Nat. Commun.* 11 (1), 2860. doi:10.1038/s41467-020-16447-7
- Tekpli, X., Lien, T., Rossevoid, A. H., Nebdal, D., Borgen, E., Ohnstad, H. O., et al. (2019). An independent poor-prognosis subtype of breast cancer defined by a distinct tumor immune microenvironment. *Nat. Commun.* 10 (1), 5499. doi:10.1038/s41467-019-13329-5
- Thompson, E. M., Keir, S. T., Venkatraman, T., Lascola, C., Yeom, K. W., Nixon, A. B., et al. (2017). The role of angiogenesis in Group 3 medulloblastoma

pathogenesis and survival. *Neuro. Oncol.* 19 (9), 1217–1227. doi:10.1093/neuonc/nox033

Tran, A. N., Dussaq, A. M., Kennell, T. J., Willey, C. D., and Hjelmeland, A. B. (2019). HPAanalyze: An R package that facilitates the retrieval and analysis of the Human Protein Atlas data. *BMC Bioinforma.* 20 (1), 463. doi:10.1186/s12859-019-3059-z

Vo, B. T., Wolf, E., Kawauchi, D., Gebhardt, A., Rehg, J. E., Finkelstein, D., et al. (2016). The interaction of myc with miz1 defines medulloblastoma subgroup identity. *Cancer Cell* 29 (1), 5–16. doi:10.1016/j.ccell.2015.12.003

Wang, S. Z., Poore, B., Alt, J., Price, A., Allen, S. J., Hanaford, A. R., et al. (2019). Unbiased metabolic profiling predicts sensitivity of high MYC-Expressing atypical Teratoid/Rhabdoid tumors to glutamine inhibition with 6-Diazo-5-Oxo-L-Norleucine. *Clin. Cancer Res.* 25 (19), 5925–5936. doi:10.1158/1078-0432.CCR-19-0189

Wang, W., Kryczek, I., Dostal, L., Lin, H., Tan, L., Zhao, L., et al. (2016). Effector T cells abrogate Stroma-Mediated chemoresistance in ovarian cancer. *Cell* 165 (5), 1092–1105. doi:10.1016/j.cell.2016.04.009

Xu-Monette, Z. Y., Deng, Q., Manyam, G. C., Tzankov, A., Li, L., Xia, Y., et al. (2016). Clinical and biologic significance of MYC genetic mutations in de novo diffuse large B-cell lymphoma. *Clin. Cancer Res.* 22 (14), 3593–3605. doi:10.1158/1078-0432.CCR-15-2296

Yamauchi, M., Yamaguchi, R., Nakata, A., Kohno, T., Nagasaki, M., Shimamura, T., et al. (2012). Epidermal growth factor receptor tyrosine kinase defines critical prognostic genes of stage I lung adenocarcinoma. *PLoS One* 7 (9), e43923. doi:10.1371/journal.pone.0043923

Yuan, J., Levitin, H. M., Frattini, V., Bush, E. C., Boyett, D. M., Samanamud, J., et al. (2018). Single-cell transcriptome analysis of lineage diversity in high-grade glioma. *Genome Med.* 10 (1), 57. doi:10.1186/s13073-018-0567-9

Zhang, J., Song, C., Tian, Y., and Yang, X. (2021). Single-Cell RNA sequencing in lung cancer: Revealing phenotype shaping of stromal cells in the microenvironment. *Front. Immunol.* 12, 802080. doi:10.3389/fimmu.2021.802080

Zhang, T., Guo, L., Creighton, C. J., Lu, Q., Gibbons, D. L., Yi, E. S., et al. (2016). A genetic cell context-dependent role for ZEB1 in lung cancer. *Nat. Commun.* 7, 12231. doi:10.1038/ncomms12231

Zhang, T., Su, F., Lu, Y. B., Ling, X. L., Dai, H. Y., Yang, T. N., et al. (2022). MYC/MAX-Activated LINC00958 promotes lung adenocarcinoma by oncogenic transcriptional reprogramming through HOXA1 activation. *Front. Oncol.* 12, 807507. doi:10.3389/fonc.2022.807507

Zhao, Y., Ma, T., and Zou, D. (2021). Identification of unique transcriptomic signatures and hub genes through RNA sequencing and integrated WGCNA and PPI network analysis in nonerosive reflux disease. *J. Inflamm. Res.* 14, 6143–6156. doi:10.2147/JIR.S340452