Genomic and immunogenomic analysis of three prognostic signature genes in LUAD

Haiming Feng
The second hospital of Lanzhou University, Lanzhou University
Ye Zhao
Gansu Provincial People's Hospital
Weijian Yan
The second hospital of Lanzhou University, Lanzhou University
Bin Li (dr.leebin@outlook.com)
The second hospital of Lanzhou University, Lanzhou University

Research Article

Keywords: lung adenocarcinoma, immunotherapy, prognostic analysis, tumor microenvironment

Posted Date: August 3rd, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1908237/v1

Additional Declarations: No competing interests reported.

Version of Record: A version of this preprint was published at BMC Bioinformatics on January 17th, 2023. See the published version at https://doi.org/10.1186/s12859-023-05137-y.

Abstract

Background

Searching for immunotherapy- related markers is an important research content to screen for target populations suitable for immunotherapy. Prognosis-related genes in early stage lung cancer may also affect the tumor immune microenvironment, which in turn affects immunotherapy.

Results

We analyzed the differential genes affecting lung cancer patients receiving immunotherapy through the Cancer Treatment Response gene signature DataBase (CTR-DB), and set a threshold to obtain a total of 176 differential genes between response and non-response to immunotherapy. Functional enrichment analysis found that these differential genes were mainly involved in immune regulation-related pathways. The early-stage lung adenocarcinoma (LUAD) prognostic model was constructed through the cancer genome atlas (TCGA) database, and three target genes (MMP12, NFE2, HOXC8) were screened to calculate the risk score of early-stage LUAD. The receiver operating characteristic (ROC) curve indicated that the model had good prognostic value, and the validation set (GSE50081) from the gene expression omnibus (GEO) analysis indicated that the model had good stability, and the risk score was correlated with immune infiltrations to varying degrees. Multi-type survival analysis and immune infiltration analysis revealed that the transcriptome, methylation and the copy number variation (CNV) levels of the three genes were correlated with patient prognosis and some tumor microenvironment (TME) components. Drug sensitivity analysis found that the three genes may affect some anti-tumor drugs. The mRNA expression of immune checkpoint-related genes showed significant differences between the high and low group of the three genes, and there may be a mutual regulatory network between immune checkpoint-related genes and target genes. Tumor immune dysfunction and exclusion (TIDE) analysis found that three genes were associated with immunotherapy response and maybe the potential predictors to immunotherapy, consistent with the CTR-DB database analysis.

Conclusions

From the perspective of data mining, this study suggests that MMP12, NFE2, and HOXC8 may be involved in tumor immune regulation and affect immunotherapy. They are expected to become markers of immunotherapy and are worthy of further experimental research.

Introduction

Lung cancer is the malignant tumor with the second highest incidence and the highest mortality in the world [1, 2]. According to the GLOBOCAN analysis report of the global tumor epidemiological statistics in 2020, the number of new cases of lung cancer worldwide reached 2.207 million, second only to breast

cancer; the number of deaths reached 1.796 million, ranking first among all cancer types. LUAD is the most common pathological type of non-small cell lung cancer (NSCLC). For driver gene-negative advanced NSCLC, the median progression free survival of traditional platinum-based doublet chemotherapy is only 4–6 months, and the median overall survival is only 10–12 months [3], and immunotherapy can bring survival benefit to driver gene-negative advanced NSCLC. Researchers [4] predicted that the advent of immunotherapy will further improve the survival outcomes of lung cancer patients, especially for advanced NSCLC with negative driver gene mutations. The food and drug administration (FDA) approved the first immune checkpoint inhibitors (ICIs) for the treatment of lung cancer in 2015. Over the past few years, the number of ICIs approved and applied in the clinic has gradually increased, and a few other anti-PD-(L)1 ICIs are currently in clinical development [5], and peptides and small peptides targeting PD-L1 have also been designed. molecules whose purpose is to block checkpoints and activate T-cell-based immunotherapy [6]. Positive responses to immunotherapy often rely on the interaction of tumor cells with immune regulation within the TME. The tumor microenvironment plays an important role in suppressing or enhancing immune responses. Understanding the interaction between immunotherapy and TME is not only the key to dissect the mechanism of action, but also of great significance to provide new methods for improving the efficacy of current immunotherapy [7, 8]. Although ICIs have shown excellent efficacy in NSCLC, their efficacy varies widely, only a subset of patients, especially those with high PD-L1 expression, benefit from long-term responses, and a large proportion of patients do not show obvious curative effect or drug resistance. For these reasons, it is necessary to combine the gene landscape of tumor immunotherapy to discover and search for potential molecules and mechanisms affecting immunotherapy, to screen target populations, and to guide individualized treatment. Obviously, even if no intervention is given after surgery for early-stage lung cancer, a good survival benefit can still be obtained. This is not only related to the biological characteristics of the tumor, but also the immune function may play a huge role in preventing tumor recurrence or distant metastasis. Therefore, we tried to find differential genes that may affect the response to immunotherapy, construct target genes that have a significant impact on the prognosis of early-stage lung cancer, and then analyze the relationship between the multi-omics changes of these genes and the tumor microenvironment of all stages of LUAD.

Materials And Methods

Immunotherapy response differential genes (ImTRDG)

The overall process of the article is shown in Fig. 1.

The mRNA and clinical data of NSCLC patients treated with anti-PD-1/PD-L1 were collected through the CTR-DB (http://ctrdb.cloudna.cn/home) [9] website, including GSE135222 [10] and GSE126044 [11] data sets from the GEO database, according to the response to immunotherapy, they were divided into responder [CR (complete response) and PR (partial response)] and non-responder [SD (stable disease) and PD(progressive disease)], and the differences in mRNA expression between the two groups were analyzed and compared. Differential genes were screened by setting the threshold adjusted P-value < =

0.05 and |logFC|>=2, and the target genes were determined according to the AUC (Area under roc Curve) value > = 0.7.

ImTRDG functional enrichment analysisc

The ImTRDG was imported into the Metascape website (https://metascape.org/) [12] for functional enrichment analysis and protein interaction analysis, and the Molecular Complex Detection (MOCODE)algorithm was used to find dense PPI MOCODE (protein-protein interaction) components in the network and annotate them. In the network analysis, set min connection to 3, p value cutoff to 0.01, and min enrichment to 1.5. In the protein interaction network analysis, the reference database is PHYSICAL_CORE, the min network size is 3, and the max network size is 500.

Univariate Cox regression analysis of ImTRDG

RNAseq data (FPKM; Fragments Per Kilobase of transcript per Million mapped reads) and corresponding clinical information for T1N0M0 stage LUAD were obtained from TCGA dataset (https://portal.gdc.com). The log-rank was used to test the Kaplan-Meier survival analysis to compare the difference in survival between the high and low expression group of ImTRDG genes. For KM curves, p-values and hazard ratios (HR) with 95% confidence intervals (CI) were derived by log rank test and univariate cox regression. $p < 0.05$ was considered statistically significant.

Prognosis signature establishment and immune infiltration nalysis

After obtaining prognostic genes through univariate cox regression, First, perform iterative analysis through multi-factor cox regression analysis, and then select the optimal model to reduce dimensionality and build a prognostic model through the step function, The model is a risk-score formula containing multiple genes, each gene has a weight, a negative number means the gene is a protective gene, and a positive number means the gene is a risk gene, and the R software glmnet package was used for the above analysis. For Kaplan-Meier curves, p-values and HR with 95% CI were obtained by log-rank test and univariate cox regression and time-ROC analysis was used to discriminate the accuracy of the prediction model. $p < 0.05$ was considered statistically significant. Finally, the stability of the model was verified using the GSE50081 dataset which is derived from the GEO database and contains the expression profiles and clinical data of 37 T1N0M0 LUAD samples[13]. Then the immune infiltration scores of T1N0M0 LUAD samples were calculated by MCpcounter method, and the correlation between risk-score and individual immune infiltration component scores was analyzed. Spearman's correlation analysis was used to describe correlations between quantitative variables without a normal distribution. p-value less than 0.05 was considered statistically significant.

Prognostic models based on gene expression and clinical characteristics

After screening the characteristic genes of T1N0M01 LUAD by cox step model, combined with clinical characteristics, firstly, univariate and multivariate cox regression analysis. Each variable (P-value, HR and 95% CI) was displayed using a forest plot by the "forestplot" package. Based on the results of a

multivariate cox proportional hazards analysis, a nomogram was built using the "rms" package to predict the year total recurrence rate. The nomogram provides a graphical result of these factors, and the prognostic time risk of an individual patient can be calculated by the points associated with each risk factor.

Expression and compiled scores Analysis

Difference analysis

LUAD gene expression data were obtained from the TCGA database and GTEx database. Based on normalized RSEM (RNA-Seq by expectation maximization) mRNA expression, fold change was calculated by mean (tumor)/mean (normal), p-value was estimated by Wilcox tests and false discovery rate (FDR) was used to analyze differences between LUAD patients in whole samples. At the same time, the Human Protein Atlas (HPA)database (https://www.proteinatlas.org/) was used to search the immunohistochemical results of the target gene translation protein in LUAD and normal samples.

Survival prognostic analysis

Merged mRNA expression and clinical survival data by sample barcode, median mRNA value was used to divide tumor samples into high and low expression groups. Then, we use R package survival to fit survival time and survival status within two groups. Cox proportional-hazards model and log rank tests were performed for every gene in LUAD. Survival types including overall survival (OS), progression free survival (PFS), disease specific survival (DSS), and disease free interval (DFI).

Potential effects of gene mRNA on pathway activity

Reverse phase protein array (RPPA) data from (The Cancer Proteome Atlas database) were used to calculate pathway activity score for TCGA LUAD samples. RPPA is a high-throughput antibody-based technique with the procedures similar to that of western blots. Proteins are extracted from tumor tissue or cultured cells, denatured by SDS, printed on nitrocellulose-coated slides followed by antibody probe. Expression and pathway activity can estimate the difference of genes expression between pathway activity groups (activation and inhibition), which defined by median pathway scores. The Gene Set Cancer Analysis (GSCA) [14] pathway included are: TSC/mTOR, RTK, RAS/MAPK, PI3K/AKT, Hormone ER, Hormone AR, EMT, DNA Damage Response, Cell Cycle, Apoptosis pathways. They are all famous cancer related pathways. RPPA data were median-centered and normalized by standard deviation across all samples for each component to obtain the relative protein level. The pathway score is then the sum of the relative protein level of all positive regulatory components minus that of negative regulatory components in a particular pathway. Samples were divided into 2 groups (high and low) by median gene expression, the difference of pathway activity score (PAS) between groups is defined by student t test, p value was adjusted by FDR, FDR < = 0.05 is considered as significant. When PAS (Gene A High expression) > PAS (Gene A Low expression), we consider gene A may have an activate effect to a pathway, otherwise have an inhibitory effect to a pathway. In addition, according to the ssGSEA (single sample gene set

enrichment analys) algorithm, the enrichment score of each sample on each pathway is calculated in turn, so as to obtain the relationship between the sample and the pathway. By calculating the correlation between the gene expression and the pathway score, the relationship between the gene and the pathway can be obtained [15, 16].

CNV and methylation analysis of target genes

CNV data of LUAD samples were downloaded from TCGA database, and were processed through GISTICS2.0, which attempts to identify significantly altered regions of amplification or deletion across sets of patients. According to the GISTIC score derived from GISTIC, CNV was classified into homozygous deletion, heterozygous deletion, heterozygous amplification and homozygous amplification. The mRNA expression data and CNV raw data were merged by TCGA barcode. CNV data and clinical survival data were merged by sample barcode. The samples were divided into WT, Amp. and Dele. groups. R survival package was used to fit survival time and survival status within groups. Log rank tests were performed to test the survival difference between groups. Finally, we integrate the CNV of a single target gene and call it a gene set CNV, the gene set CNV represents the integrated CNV status of target gene set for each sample. A sample is classified into Amp. or Dele. group. If all genes in inputted gene set have no CNV in a sample, this sample is classified into WT group. The association of gene set CNVs with survival prognosis was then analyzed.

LUAD Illumina Human Methylation 450k data were downloaded from TCGA database, Methylation data and clinical survival data were combined by sample barcodes. The median methylation was used to classify tumor samples into hypermethylated and hypomethylated groups. The cox proportional-hazards model was constructed to obtain the hazard ratio of the hypermethylated group to the hypomethylated group. A log rank test was performed to test whether the difference in survival between groups was statistically significant.

Immune infiltration analysis

The infiltration of 24 immune cells was assessed by ImmuCellAI, and the association between gene mRNA expression, gene CNVs (copy number variations), gene methylation and gene set CNVs and immune cell infiltration was estimated [17, 18].

Drug Sensitivity Analysis

We collected the IC50s and their corresponding mRNA gene expressions of 481 small molecules in 1001 cell lines from the Genomics of Therapeutics Response Portal (CTRP) [19]. Also Genomics of Drug Sensitivity in Cancer (GDSC) [20] contained the IC50 of 265 small molecules in 860 cell lines, the IC50 corresponding mRNA gene expression from mRNA expression data and drug sensitivity data were combined. Pearson correlation analysis was performed to obtain the correlation between gene mRNA expression and drug IC50. At the same time, the correlation between the expression of target gene and

related drugs IC50 was analyzed by Consortium for Classical Lutheran Education (CCLE) (http://www.ccle.org/) drug response database.

Expression and network relationship between target genes and immune checkpoint genes

RNAseq data and corresponding clinical information for LUAD were obtained from TCGA dataset. SIGLEC15, TIGIT, CD274, HAVCR2, PDCD1, CTLA4, LAG3 and PDCD1LG2 are genes related to immune checkpoints. The expression values of these 8 genes were extracted to observe the expression of target genes related to immune checkpoints. According to the differential relationship between target genes and immune checkpoint genes, use the Gene Network Search function on the GenCLiP 3 website (http://ci.smu.edu.cn/genclip3/ analysis.php) [21] to search for target genes and immune checkpoint genes with significant differences interaction networks and analyze possible regulatory relationships.

Analysis of target gene and immune efficacy

The TCGA LUAD gene expression data were obtained, and the TIDE algorithm [22,23] was used to predict the response of the high and low expression groups of the target gene to the predicted immune checkpoint inhibitor. TIDE uses a panel of gene expression signatures to assess 2 distinct tumor immune escape mechanisms, including tumor-infiltrating cytotoxic T lymphocyte (CTL) dysfunction and CTL rejection by immunosuppressive factors. High TIDE score, poor response to immune checkpoint blockade (ICB), and short survival after receiving ICB.

Results

Identity of ImTRDG and functional enrichment results

The two datasets GSE135222 and GSE126044 in the CTR-DB database are about NSCLC patients who received immunochemotherapy, with a total of 43 patients. According to the effect of immunotherapy, they were divided into 13 responders (CR and PR) and 30 non-responders (SD and PD), as shown in Table 1. Through differential analysis and the set threshold, a total of 176 differential genes were screened, of which 72 were up-regulated genes and 104 genes were down-regulated. Figures 2A and Figs. 2B showed the volcano map and heat map of differential genes.

Table 1
Overview of receiving immunotherapy dataset information. complete response (CR), partial response (PR), stable disease (SD), progressive disease (PD).

Treatment response	Number	Data sets	Medication regimen	Pathological type
non-responder (SD and PD)	19	GEO:GSE135222	Immunotherapy	Lung non-small cell carcinoma
responder (CR and PR)	8	GEO:GSE135222	anti-PD-1/PD-L1	Lung non-small cell carcinoma
non-responder (SD and PD)	6	GEO:GSE126044	Nivolumab	Lung squamous cell carcinoma
non-responder (SD and PD)	5	GEO:GSE126044	Nivolumab	LUAD
responder (CR and PR)	3	GEO:GSE126044	Nivolumab	Lung squamous cell carcinoma
responder (CR and PR)	2	GEO:GSE126044	Nivolumab	LUAD

Through metascape enrichment analysis, the top 20 significant results were extracted, suggesting that most of the differential genes are involved in immune-related pathways (Fig. 3A), through the MOCODE algorithm, we obtained three densely connected PPI (protein-protein interaction) MCODE components, which are involved in neutrophil degranulation, regulation of natural killer cell mediated cytotoxicity and CD8 TCR (T cell receptor) pathway, respectively, as shown in Fig. 3B.

Target genes for model screening and immune infiltration nalysis

After obtaining 176 lmTRDGs, univariate cox regression analysis was performed, and a total of 4 genes (CLEC4E, HOXC8, MMP12 and NFE2) were obtained that were associated with the prognosis of T1N0M0 LUAD as shown in Fig. 4A-D. Through multivariate cox and step functions, the risk model (Riskscore= -1.0068*NFE2 + 0.2741*MMP12 + 0.5986*HOXC8) constructed by 3 genes (HOXC8, MMP12 and NFE2), 81 samples can be divided into high and low groups according to riskscore, survival analysis showed that the survival difference between the high-risk group and the low-risk group was statistically significant (HR = 3.491, 95%CI: 1.062–11.475, P = 0.0395). The 1-year, 3-year and 5-year ROC curve area was 0.916, 0.90 and 86.3, respectively (Fig. 5A-D). The GSE50081 data was used to verify the accuracy of the model. The results showed that the survival prognosis of patients in the high and low risk groups was statistically different (p = 0.048) (Fig. 6A), and the 5-year ROC curve area was 0.67(Fig. 6B), which was relatively stable. Combined with clinical data (age, gender and smoking status), univariate and multivariate cox regression analysis was performed, it has showed that MMP12, NFE2 and HOXC8 can be used as independent prognostic factors for T1N0M0 LUAD (Supplementary Fig. 1A-B). The correlation analysis between riskscore and immune infiltration scores showed that there was a good positive correlation between riskscore and cytotoxicity, NK.cell and CD8T cell scores (Fig. 6C).

Differential expression and multi-type prognostic analysis

The differential expression of the three genes in LUAD and normal samples from GTEx database was analyzed, and the results showed that MMP12 and HOXC8 were highly expressed in LUAD samples, while NFE2 was low expressed in LUAD samples (Supplementary Fig. 2A-C). The immunohistochemical information of NFE2 in LUAD and normal samples was searched by HPA, using HPA001914 antibody labeling, the results indicated that NFE2 was moderately stained in normal alveolar tissue, but low in LUAD. Using HPA028911 antibody labeling, it was found that HOXC8 was not stained in normal alveolar tissue, but moderately stained in alveolar macrophages, and HOXC8 was moderately stained in LUAD tumor tissue. The results of immunohistochemistry and mRNA expression were consistent. (Fig. 7A-D) Survival analysis showed that FOXC8 high expression was significantly correlated with poor OS, PFS, DSS, and DF of LUAD (Fig. 8A-D).

Gene expression and pathway activity result

GSCA–Expression and pathway activity module estimated difference of three genes expression between pathway activity groups (activation and inhibition). The results showed that NFE2 may have inhibitory effects on the Apoptosis, CellCycle, EMT and Hormone AR pathways of LUAD, while it has an activation effect on the MAPK and mTOR pathways. MMP12 has an activating effect on Apoptosis, CellCycle and EMT pathways of LUAD, and has an inhibitory effect on MAPK pathway, and HOXC8 has an activating effect on CellCycle pathway (Fig. 9A). Through pathway ssGSEA analysis, the relationship between target genes and pathway scores was calculated and found that MMP12 had positive correlation with Cellular_response_to_hypoxia, Tumor_proliferation_signature, G2M_checkpoint, Tumor_Inflammation_Signature and DNA_repair. NFE2 has negative correlation with Tumor_proliferation_signature and G2M_checkpoint, and HOXC8 has positive correlation with Tumor_proliferation_signature and G2M_checkpoint (Fig. 9B).

Copy Number Variation (CNV) and methylation survival prognostic analysis of target genes

The summary of CNV of target genes in LUAD shown in the Table 2. The results of the CNV and LUAD survival prognostic analysis showed that compared with the WT group, the HOXC8 and NFE2 CNV groups were associated with poor OS (Fig. 10A-B), and NFE2 CNV groups were also associated with poor PFS in LUAD (Fig. 10C). The MMP12 CNV was associated with poor DFI (Fig. 10D). The results of CNV and survival prognostic analysis after the integration of the three genes showed that the gene set CNV was associated with poor OS in LUAD(Fig. 10E). Survival analysis showed that MMP12 hypermethylation levels were associated with good DFS, DSS and DFI (Fig. 9F-H), and HOXC8 hypermethylation levels were associated with poor PFS and DFI in LUAD (Fig. 10I-J).

Table 2
The summary of CNV of target genes in LUAD. Total amp.(%): the percentage of samples with copy number amplification, including heterozygous and homozygous amplification; Total dele.(%): the percentage of samples with copy number deletion, including heterozygous and homozygous deletion;

Symbol	Total amp. (%)	Total dele. (%)	Hete amp. (%)	Hete dele. (%)	Homo amp. (%)	Homo dele. (%)
HOXC8	28.68217	18.21705	27.51938	18.02326	1.162791	0.193798
MMP12	26.16279	19.37985	24.22481	18.99225	1.937985	0.387597
NFE2	28.48837	18.60465	27.32558	18.41085	1.162791	0.193798

Hete amp.(%): the percentage of samples with copy number heterozygous amplification; Hete dele.(%): the percentage of samples with copy number heterozygous deletion; Homo amp.(%): the percentage of samples with copy number homozygous amplification; Homo dele.(%): the percentage of samples with copy number homozygous deletion.

Drug Sensitivity Analysis

From the GDSC database, we analyzed that NFE2 mRNA expression was correlated with IC50 of Nilotinib, TL-1-85 and BHG712, and MMP12 mRNA expression was negatively correlated with Gefitinib IC50(Fig. 11A); CTRP database analysis found that NFE2 mRNA expression was negatively correlated with BRD-K01737880 IC50, HOXC8 mRNA expression was positively correlated with tacedinaline, JQ-1 IC50 (Fig. 11B); CCLE database results indicated that the FGFR targeting drug TKI258 IC50 difference was statistically significant in the HOXC8 mRNA high and low expression groups, In the MMP12 mRNA high and low expression groups, the IC50 differences of c-MET targeting drug PF2341066, ALK targeting drug TAE684 and IGF1R targeting drug AEW541 were statistically significant (Supplementary Fig. 3A-B).

Immune infiltration analysis

Gene expression and immune infiltration analysis showed that MMP12 was correlated with many immune infiltration components, among which it was positively correlated with nTreg, iTreg and Exhausted, and negatively correlated with Th17 and Th2. NFE2 expression was negatively correlated with central_memory, HOXC8 expression was positively correlated with nTreg, and negatively correlated with Gamma_delta and MAIT (Mucosal Associated Invariant T) (Fig. 12A).

The results of gene CNV and immune infiltration showed that NFE2 and HOXC8 CNV were positively correlated with nTreg and negatively correlated with CD4_T and Th2(Fig. 12B).

The results of gene methylation and immune infiltration analysis showed that MMP12 methylation was negatively correlated with ntreg cells and positively correlated with CD_4T cells, NFE2 methylation was positively correlated with Th17, and negatively correlated with NK cells, Th1 cells, Cytotoxic and Exhausted cells, and HOXC8 methylation was positively correlated with DCs. cells, CD_4T cells were positively correlated (Fig. 12C).

After integrating the CNV results of the three genes, their relationship with immune infiltration was analyzed, and it was found that nTreg, exhausted, effector_memory, monocyte, neutrophil, Th1 cells aggregated in high CNV tumors, while CD4_naive, Th2, Tfh, NKT (Natural killer T cell), Gamma_delta, NK, MAIT and CD4_T aggregated in low CNV tumors (Fig. 12D).

Expression Relationship and network between target genes and immune checkpoint genes

The correlation analysis of the three genes and immune checkpoint genes found that only MMP12 had a weak linear correlation with immune checkpoint genes (Fig. 13A). The target genes were divided into high and low expression groups according to their expression levels. Between the high and low expression groups of HOXC8, the expressions of CD274 and HAVCR2 were significantly different (Fig. 13B). The expressions of HAVCR2, PDCD1LG2, CTLA4, TIGIT, LAG3 and PDCD1 were all different in the NFE2 high and low expression groups (Fig. 13C). In the high and low expression groups of MMP12, the expressions of SIGLEC15, TIGIT, CD274, HAVCR2, PDCD1, CTLA4, LAG3 and PDCD1LG2 were statistically different (Fig. 13D). Through the GenCLiP 3 website to analyze the potential regulatory networks of risk target genes and immune check genes, some of them have been confirmed by experiments, and some regulatory networks still need to be verified in the experimental area, as shown in the Fig. 14.

Analysis of target gene and immune efficacy

The TIDE algorithm was used to calculate the response of the high and low expression LUAD of the three target genes to immunotherapy (Table 3). The results showed that 110 patients in the NFE2 high expression group responded to immunotherapy, 147 patients did not respond to immunotherapy, 87 patients in the NFE2 low expression group responded to immunotherapy, and 169 patients did not respond to immunotherapy. The TIDE score results showed that the TIDE score of the NFE2 low expression group was higher, indicating that the effect of immunotherapy was poor, indicating that the high expression of NFE2 may be a positive indicator of immunotherapy (Fig. 15A). This is consistent with the CTR-DB immunotherapy response differential gene results (Fig. 15B).

Table 3
Statistical table of immune responses of samples in different groups in prediction results.

Response	NFE2		HOXC8		MMP12	
	high	low	high	low	high	low
True	110	87	85	112	119	78
False	147	169	172	144	137	179

In the HOXC8 high expression group, 85 patients responded to immunotherapy, 172 patients did not respond to immunotherapy, 112 patients in the HOXC8 low expression group responded to immunotherapy, and 144 patients did not respond to immunotherapy. The TIDE score results showed that

the TIDE score was higher in the high HOXC8 expression group, indicating that the immunotherapy effect was poor, which means that the high expression of HOXC8 may be a negative indicator of immunotherapy (Fig. 15C), which is consistent with the CTR-DB immunotherapy response differential gene results (Fig. 15D).

In the MMP12 high expression group, 119 patients responded to immunotherapy, 137patients did not respond to immunotherapy, 78 patients in the MMP12 low expression group responded to immunotherapy, and 179 patients did not respond to immunotherapy. The TIDE score results showed that the TIDE score of the MMP12 low expression group was higher, indicating that the effect of immunotherapy was poor, indicating that the high expression of MMP12 may be a positive indicator of immunotherapy (Fig. 15E). This is consistent with the CTR-DB immunotherapy response differential gene results (Fig. 15F).

Discussion

With the advent of immunotherapy in recent years, the treatment and natural history of advanced NSCLC has been revolutionized, and immunotherapy for squamous cell carcinoma appears to yield better results than adenocarcinoma. In fact, in patients with driver-negative LUAD, the benefit of immune checkpoint inhibitors (ICIs) over previous standard chemotherapy has been demonstrated in first-line and further first-line therapy [23–25]. However, despite the overall benefit in survival outcomes, a large proportion of NSCLC patients were observed to experience disease progression. Exactly why this difference occurs and how to predict the effect of immunotherapy is still an important part of the ongoing research in the field of immunotherapy. Scientists have made great efforts to evaluate predictive biomarkers [26]. So far, only the high expression of programmed death ligand-1 (PD-L1) demonstrated by immunohistochemistry has been confirmed for screening target populations even in different treatment stages and different immunotherapy regimens of LUAD predictive biomarkers. TMB (tumor mutational burden)/ bTMB (blood tumor mutational burden) has also been regarded as a predictor of immunotherapy. However, current studies have shown that TMB/bTMB as a predictor of ICIs treatment effect is still controversial. Exploratory analyses of CheckMate-026 [27] and POPLAR [28]/OAK [29] studies suggest that patients with high TMB/bTMB can benefit from immunotherapy, while the results of an exploratory analysis of the KEYNOTE series showed that TMB was not associated with efficacy, regardless of whether TMB was high or low, pembrolizumab plus chemotherapy in the first-line treatment of both squamous and non-squamous NSCLC patient survival benefit [30]. It has been reported that the clinical application of pembrolizumab in the treatment of advanced tumors was guided and the clinical efficacy of pembrolizumab was predicted based on the expression level of mismatch repair (MMR) [31]. The CheckMate-142 clinical study evaluated the efficacy of nivolumab monotherapy versus nivolumab in combination with ipilimumab in the treatment of metastatic colorectal cancer, in MSI-H colorectal cancer patients, ORR was better in both monotherapy and combination therapy groups than in patients with stable microsatellites. [32]. Although MMR status may be used to predict the efficacy of PD-1/PD-L1 inhibitors, due to its low incidence in lung cancer, the predictive value of dMMR/MSI-H for lung cancer immunotherapy efficacy needs more research data to verify.

The TME is composed of tumor cells, stromal cells (including vascular endothelial cells, pericytes, immune inflammatory cells, etc.) and extracellular matrix. The TME is not only the basis of tumor growth, invasion and metastasis [33], but also affects the clinical treatment effect of various cancers [34]. The tumor microenvironment has gradually become a research hotspot in recent years. Studies have shown that the interaction between cancer cells and the TME is bidirectional and dynamic, and the microenvironment has both promotion and inhibition on the occurrence and development of tumors. Like other malignant tumors, lung cancer is infiltrated with a large number of immune cells around the tumor, mainly T cells, macrophages and mast cells, while the relative content of plasma cells, natural killer cells and myeloid suppressor cells is relatively low [35,36]. However, the specific cell composition has certain heterogeneity according to different tumor subtypes and patients [34]. The type, density, location and function of immune cells together constitute a specific immune context [37]. A large number of studies have shown that lymphocytes infiltrated by in situ tumors and metastases are closely related to tumor development and clinical outcomes of patients [38, 39]. The density of different cells in the immune microenvironment has a certain correlation with the survival of NSCLC, and has a strong prognostic value [34, 40].

We screened differential genes in response to immunotherapy, and functional enrichment analysis found that target genes are mainly involved in the process of immune stripping. Using TCGA data to build a prognostic model for early-stage LUAD, the model constructed from three genes has good predictive value. The immune infiltration of individual genes of interest can also be analyzed in all stages of LUAD. The predicted AUC values at 1, 3, and 5 years were 0.916 (95%CI: 0.859–0.973), 0.9 (95%CI: 0.809–0.99), and 0.863 (95%CI: 0.724–1.002), respectively. Pathway activity analysis found that three genes were involved in EMT, tumor proliferation, cell cycle cycle, cell damage repair, MAPK and mTOR pathway to varying degrees.

Homeobox C8 (HOXC8) belongs to the HOX family, comprising 39 members in mammals, and the HOXC8 protein is involved in many physiological and pathological processes, including embryogenesis and tumorigenesisv[41]. HOXC8 has been reported to be dysregulated in various types of cancer, including breast, cervical, prostate, and ovarian cancer, and acts as a transcription factor to regulate the transcription of many genes [42]. HOXC8 was significantly upregulated in NSCLC clinical specimens compared with normal tissues which is consistent with our TCGA database analysis results. And the upregulation of HOXC8 played an important role in the tumorigenicity of NSCLC cell lines A549 and NCI-H460[43]. Loss of E-cadherin expression is a hallmark of epithelial-mesenchymal transition (EMT) in tumor progression. Liu et al [43] found that HOXC8 could promote EMT in NSCLC, and E-cadherin was the target gene of HOXC8, the loss of E-cadherin promoted the growth and migration of NSCLC. The results of our pathway ssGSEA analysis also showed that HOXC8 had a weak linear relationship with EMT pathway scores (Pearce correlation coefficient is 0.22, $p < 0.05$). Yu et al [44]. found that HOXC8 is a key biomarker for glioma diagnosis and prognosis through biological information, and the expression level of HOXCs is related to the infiltration of various immune cells. The prognostic value of HOXC8 in glioma was further validated by qPCR and immunohistochemical data. The results of our immune infiltration analysis showed that HOXC8 mRNA expression had a weak positive linear correlation with nTreg cell, and

a weak negative linear correlation with Gamma_delta and MAIT cell. The results of gene CNV and immune infiltration showed that HOXC8 CNV were positively correlated with nTreg and negatively correlated with CD4_T and Th2. And the results of gene methylation and immune infiltration analysis showed that HOXC8 was positively correlated with DCs. cells, CD_4T cells. The correlation analysis between target genes and immune checkpoints showed that the expressions of CD274 and HAVCR2 were significantly different between high and low expression groups of HOXC8. TIDE analysis suggested that HOXC8 may be a negative indicator of immunotherapy, which was basically consistent with the results of immune infiltration analysis. Although there is no strong linear relationship between HOXC8 and immune checkpoint-related genes, the GenCLiP 3 website analysis found that HOXC8 may have a complex regulatory network with immune checkpoint-related genes. In addition, drug sensitivity analysis found that HOXC8 may affect the antitumor effect of multiple drugs. Experiments related to HOXC8 methylation, CNV and immune infiltration of LUAD are still blank, and further basic experiments need to be carried out to prove it.

NFE2 (Nuclear Factor, Erythroid 2) is a Protein Coding gene. Diseases associated with NFE2 include Erythroleukemia and Polycythemia. Among its related pathways are Response to elevated platelet cytosolic Ca2 + and Hematopoietic Stem Cell Differentiation [45, 46]. There are few reports on the relationship between NFE2 and tumors. Wang et al [47]. analyzed lung cancer transcriptome sequencing and genomic data and found a novel R3HDM2-NFE2 fusion in the H1792 lung cancer cell line. Lung tissue microarray revealed that 2 of 76 lung cancer patients had genomic rearrangements at the NFE2 locus, and when NFE2 was knocked down, it reduced the proliferation and invasion of H1792 cells. Dou et al [48]. found that NFE2 members bind to the antioxidant response element region and activate the expression of target genes. Through bioinformatics analysis, they showed that NFE2 members mainly focus on transcriptional coactivator activities. The mRNA expression of NFE2 members was significantly correlated with the immune infiltration of CD4 + T cells, CD8 + T cells, B cells, macrophages and neutrophils in Ovarian Cancer. The results of our immune infiltration analysis showed that NFE2 expression was negatively correlated with Central_memory. Central memory T cells which are restricted to the secondary lymphoid tissues and blood are with long-term memory generated after naive T cells are activated by antigens, and can home to lymph nodes to receive antigen re-stimulation. Continue to generate large numbers of alloantigen-bearing cloned effector memory T cells upon restimulation. In 2005, Klebanoff CA et al. first proved that Central memory T cells have super anti-tumor ability [49]. In 2012, clinical studies such as the National Institutes of Health (NIH) found that Central memory T cells and their derived cloned T cells are highly effective anti-tumor cells. Tumor immune T cells [50]. Collecting the results of our analysis, we hypothesized that NFE2 may be associated with tumor tertiary lymph nodes and circulating tumor cells in LUAD cells. The results of gene CNV and immune infiltration showed that NFE2 CNV were positively correlated with nTreg and negatively correlated with CD4_T and Th2. And the results of gene methylation and immune infiltration analysis showed that NFE2 was positively correlated with Th17, and negatively correlated with NK cells, Th1 cells, Cytotoxic and Exhausted cells. The correlation analysis between target genes and immune checkpoints showed that the expressions of HAVCR2, PDCD1LG2, CTLA4, TIGIT, LAG3 and PDCD1 were all different in the NFE2 high

and low expression groups. TIDE analysis suggested that NFE2 may be a positive indicator of immunotherapy, which was basically consistent with the results of immune infiltration analysis. The above dry analysis results still need experiments to enhance convincing.

Matrix metalloproteinases (MMPs) are a group of more than 20 proteolytic enzymes that degrade the extracellular matrix and facilitate invasion through the basement membrane [51, 52]. This ability of MMPs to remodel the extracellular milieu has led to extensive studies of their role in carcinogenesis. In NSCLC, MMPs are implicated in tumor invasion and metastasis through their ability to remodel and degrade the extracellular matrix and mediate cell-cell adhesion [53, 54]. In addition to disrupting the basement membrane, MMPs have been shown to influence the microenvironment of cells through complex cell-cell and cell-matrix interactions, by altering cell signaling and regulating cytokines, growth factors, and angiogenic factors [55]. Hofmann et al. [56] found that MMP-12 expression was significantly increased in tumors compared with corresponding lung tissues, and MMP-12 expression was significantly associated with local recurrence and metastatic disease. Multivariate Cox regression analysis showed that MMP-12 expression was an independent prognostic factor for tumor recurrence-free interval. Immunohistology identified MMP-12 protein in NSCLC only in tumor cells. Hung et al [57]. found that nontoxic concentrations of penfluidol reduced LUAD cell migration, invasion, and adhesion. A protease array screen identifies MMP-12 as a potential target of penfluridor to modulate LUAD cell motility and adhesion. Mechanistic studies showed that penfluridol downregulates MMP-12 expression by inhibiting the urokinase plasminogen activator (uPA)/uPA receptor/transforming growth factor-beta/Akt axis, thereby reversing MMP-12-induced EMT. Subsequent analysis of clinical LUAD samples revealed a positive correlation between MMP12 and mesenchymal-related gene expression levels. In addition, some studies have found that MMP12 may be involved in the MAPK pathway to affect cell damage and repair[58,59]. These findings are consistent with our pathway activity analysis results. Regulatory T cells (Tregs) are a subset of immune cells, including nTregs and iTregs, both of which play a role in suppressing immunity and promote tumor progression by suppressing antitumor immune responses [60]. Kim et al. [61] used an anti-ST2 antibody to deplete Tregs in mouse lung tumors and found that local Tregs depletion resulted in a significant reduction in lung tumor burden. Immune responses following depletion of Tregs in tumors showed restoration of NK cell activity, enhanced Th1 activity, increased CD8 cytotoxic T cell responses, and decreased expression of Mmp12. Our immune infiltration analysis found that MMP12 showed a positive linear relationship with nTreg and iTreg, indicating that high expression of MMP12 may mean increased nTreg and iTreg, promoting tumor growth, suggesting that MMP12 may be a negative factor for immunotherapy, and our TIDE The analysis found that the higher the expression of MMP12, the higher the TIDE score and the worse the immunotherapy effect, which is consistent with the above findings. These data suggest that therapeutic strategies targeting activated Tregs in lung cancer have the potential to inhibit tumor progression by enhancing antitumor immunity. In addition, we analyzed the relationship between MMP12 methylation levels and immune infiltration and found that MMP12 methylation was negatively correlated with nTreg cells and positively correlated with CD_4T cells.The correlation analysis between target genes and immune checkpoints showed that the expressions of SIGLEC15, TIGIT, CD274, HAVCR2, PDCD1, CTLA4,

LAG3 and PDCD1LG2 were significantly different between high and low expression groups of MMP12. The GenCLiP 3 website analysis found that MMp12 may have a complex regulatory network with immune checkpoint-related genes. In addition, drug sensitivity analysis found that MMP12 may affect the antitumor effect of multiple drugs. However, the above analysis results still need accurate experimental data to verify.

Conclusions

In conclusion, our bioinformatic results suggest that the early-stage LUAD prognostic model constructed by MMP12, NFE2, and HOXC8 has good predictive value; MMP12, NFE2, and HOXC8 are involved in the formation and growth pathway of LUAD to varying degrees and may affect the The effect of some antitumor drugs; the mRNA expression, methylation level and CNV status of MMP12, NFE2, HOXC8 have a certain linear relationship with some immune infiltration components, which may be involved in the immune regulation of tumors; MMP12, NFE2, HOXC8 and immune examination Dot-related genes have complex regulatory networks that affect immunotherapy and are expected to be markers of immunotherapy, which are worthy of further experimental research. However, the bioinformatics analysis still lacks strong convincing power and needs to be verified by subsequent experiments.

Abbreviations

CTR-DB Cancer Treatment Response Gene Signature DataBase

LUAD Lung Adenocarcinoma

TCGA The Cancer Genome Atlas

MOCODE Molecular Complex Detection

MMP12 Matrix Metallopeptidase 12

NFE2 Nuclear Factor, Erythroid 2

HOXC8 Homeobox C8

ROC Receiver Operating Characteristic

GEO The Gene Expression Omnibus

CNV The Copy Number Variation

TME Tumor Microenvironment

TIDE Tumor Immune Dysfunction and Exclusion

NSCLC Non-small Cell Lung Cancer

FDA The Food and Drug Administration

ICIs Immune Checkpoint Inhibitors

PD-L1 Programmed cell death ligand 1

PD-1 Programmed cell death 1

ImTRDG Immunotherapy Response Differential Genes

CR Complete Rresponse

PR Partial Response

SD Stable Disease

PD Progressive Disease

HR Hazard Ratios

CI Confidence Intervals

FDR False Discovery Rate

HPA The Human Protein Atlas

OS Overall Survival

PFS Progression Free Survival

DSS Disease Specific Survival

DFI Disease Free Interval

RPPA Reverse Phase Protein Array

GSCA The Gene Set Cancer Analysis

PAS Pathway Activity Score

SsGSEA Single Aample Gene Set Enrichment Analys

TSC Tuberous Sclerosis

MTOR Mammalian Target of Rapamycin

RTK Receptor Tyrosine Kinases

MAPK Mitogen-Activated Protein Kinase

PI3K Phosphoinositide-3 Kinase

ER Estrogen Receptor

AR Androgen Receptor

EMT Epithelial–Mesenchymal Transition

DNA DeoxyriboNucleic Acid

mRNA Messenger Ribonucleic Acid

CTRP Genomics of Therapeutics Response Portal

GDSC Genomics of Drug Sensitivity in Cancer

SIGLEC15 Sialic Acid Binding Ig Like Lectin 15

TIGIT T cell Immunoreceptor with Ig and ITIM domains

HAVCR2 Hepatitis A virus Cellular Receptor 2

PDCD1 programmed cell death 1

CTLA4 Cytotoxic T-lymphocyte Associated Protein 4

LAG3 Lymphocyte activating 3

PDCD1LG2 Programmed cell death 1 ligand 2

CTL Cytotoxic T Lymphocyte

ICB Immune Checkpoint Blockade

PPI Protein-Protein Interaction

CLEC4E C-type Lectin Domain family 4 member E

GTEx Genotype-Tissue Expression

ALK Anaplastic Lymphoma Kinase

FGFR Fibroblast Growth Factor Receptor

CCLE Cancer Cell Line Encyclopedia

MAIT Mucosal Associated Invariant T

NKT Natural Killer T cell

TMB Tumor Mutational Burden

bTMB Blood Tumor Mutational Burden

MMR Mismatch Repair

R3HDM2 R3H domain Containing 2

MMPs Matrix Metalloproteinases

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

The datasets(GSE135222, GSE126044 and GSE50081) generated and analysed during the current study are available in the GEO dataset repository. https://www.ncbi.nlm.nih.gov/gds.The original contributions presented in the study are included in the article/Supplementary Material. Further inquiries can be directed to the corresponding authors. The Supplementary Material for this article can be found online.

Competing interests

The authors declare that they have no competing interests.

Funding

Not applicable

Authors' contributions

LB designed the study. FHM and YWJ performed the data analysis. FHM and ZY wrote the manuscript and helped with the validation. YWJ and ZY helped the revision. All authors contributed to the article and approved the submitted version.

Acknowledgements

We thank the website “Lin chuang sheng xin zhi jia” (https://www.aclbi.com/static/ index.html#/) for the data visualization support.

References

1. Siegel R L, Miller K D, Fuchs HE, Jemal A. CA Cancer J Clin. 2021;71:7-33. doi:10.3322/caac.21654.
2. Sung H, Ferlay J, Siegel R L, Laversanne M, Soerjomataram I, Jemal A, et al. Global cancer statistics 2020: GLOBOCA N estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2021;71:209-249. doi: 10.3322/caac.21660.
3. Socinski M.A, Bondarenko I, Karaseva N.A, Makhson A.M, Vynnychenko I, Okamoto I, et al. Weekly nab-paclitaxel in combination with carboplatin versus solvent-based paclitaxel plus carboplatin as first-line therapy in patients with advanced non-smallcell lung cancer: final results of a phase III trial. J Clin Oncol. 2012;30:2055-62. doi: 10.1200/JCO.2011.39.5848.
4. Howlader N, Forjaz G, Mooradian M.J, Meza R, Kong C.Y, Cronin K.A, et al. The effect of advances in lung-cancer treatment on population mortality. N Engl J Med. 2020;383:640-649. doi: 10.1056/NEJMoa1916623.
5. Akinleye A, Rasool Z. Immune checkpoint inhibitors of PD-L1 as cancer therapeutics. J Hematol Oncol. 2019;92. Doi:10.1186/s13045-019-0779-5
6. Constantinidou A, Alifieris C, Trafalis D.T, Targeting programmed cell death−1 (PD-1) and ligand (PD-L1): a new era in cancer active immunotherapy. Pharmacol Ther. 2019;194:84–106. doi: 10.1016/j.pharmthera.2018.09.008
7. Fridman W.H, Dieu-Nosjean M.C, Pagès F, Cremer I, Damotte D, Catherine SF, et al. The Immune Microenvironment of Human Tumors: General Significance and Clinical Impact. Cancer Microenvironment. 2013;6:117-22. doi: 10.1007/s12307-012-0124-9.
8. Fridman, W.H, Remark R, Goc J, Giraldo N.A, Becht E, Hammond S.A, et al. The Immune Microenvironment: A Major Player in Human Cancers. Int Arch Allergy Immunol. 2014;164:13-26. doi: 10.1159/000362332.
9. Liu Z.Y, Liu J.L, Liu X.Y, Wang X, Xie Q.S, Zhang X.L, et al. CTR-DB. an omnibus for patient-derived gene expression signatures correlated with cancer drug response. Nucleic Acids Res. 2022;50:D1184-D1199. doi: 10.1093/nar/gkab860.
10. Kim J.Y, Choi J.K, Jung H. Genome-wide methylation patterns predict clinical benefit of immunotherapy in lung cancer. Clin Epigenetics. 2020;1:119. doi: 10.1186/s13148-020-00907-4.
11. Cho J.W, Hong M.H, Ha S.J, Kim Y.J, Cho BC, Lee I, et al. Genome-wide identification of differentially methylated promoters and enhancers associated with response to anti-PD-1 therapy in non-small cell lung cancer. Exp Mol Med. 2020; 52:1550-1563. doi: 10.1186/s13148-020-00907-4.
12. Zhou Y, Zhou B, Pache L, Chang M, Khodabakhshi A.H, Tanaseichuk O, et al. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. Nature Communications.

2019;10:1523. doi: 10.1038/s41467-019-09234-6.

13. Der SD, Sykes J, Pintilie M, Zhu CQ, Strumpf D, Liu N, et al. Validation of a histology-independent prognostic gene signature for early-stage. non-small-cell lung cancer including stage IA patients. J Thorac Oncol. 2014;9:59-64. doi: 10.1097/JTO.0000000000000042.
14. Liu C.G, Hu F.F, Xia M.X, Han L, Zhang Q, Guo A.Y, et al. GSCALite: a web server for gene set cancer analysis. Bioinformatics. 2018;21:3771-3772. doi: 10.1093/bioinformatics/bty411.
15. Wei J, Huang K, Chen Z, Hu M, Bai Y, Lin S, et al. Characterization of Glycolysis-Associated Molecules in the Tumor Microenvironment Revealed by Pan-Cancer Tissues and Lung Cancer Single Cell Data. Cancers (Basel). 2020;12:1788. doi: 10.3390/cancers12071788.
16. Hänzelmann S, Castelo R, Guinney J. GSVA: gene set variation analysis for microarray and RNA-seq data. BMC Bioinformatics. 2013;14:7. doi: 10.1186/1471-2105-14-7.
17. Miao Y.R, Zhang Q, Lei Q, Luo M, Xie G.Y, Wang H.X, et al. ImmuCellAI: A Unique Method for Comprehensive T-Cell Subsets Abundance Prediction and its Application in Cancer Immunotherapy. Adv Sci. 2020; 7:1902880. doi: 10.1002/advs.201902880.
18. Miao YR, Xia MX, Luo M, Luo T, Yang M, Guo A.Y. An-Yuan Guo. ImmuCellAI-mouse: a tool for comprehensive prediction of mouse immune cell abundance and immune microenvironment depiction. Bioinformatics. 2021; btab711. doi: 10.1093/bioinformatics/btab711.
19. Basu A, Bodycombe N.E, Cheah J.H, Price E.V, Liu K, Schaefer G.I, et al. An interactive resource to identify cancer genetic and lineage dependencies targeted by small molecules. Cell. 2013;154:1151-1161. doi: 10.1016/j.cell.2013.08.003.
20. Wanjuan Y, Jorge S, Patricia G, Edelman E.J, Lightfoot H, Forbes S, et al. Genomics of Drug Sensitivity in Cancer (GDSC): a resource for therapeutic biomarker discovery in cancer cells. Nucleic Acids Research. 2013; 41:D95 -61. doi: 10.1093/nar/gks1111.
21. Wang JH, Zhao LF, Wang HF, Wen YT, Jiang KK, Mao X.M, et al. GenCLiP 3: mining human genes' functions and regulatory networks from PubMed based on co-occurrences and natural language processing. Bioinformatics. 2020;36: 1973-1975. Doi:10.1093/bioinformatics/btz807.
22. Jiang P, Gu S, Pan D, Fu J, Sahu A, Hu, H.X, et al. Signatures of T cell dysfunction and exclusion predict cancer immunotherapy response. Nat Med. 2018;24:1550–1558. doi: 10.1038/s41591-018-0136-1
23. Wang Q, Li M, Yang M, Yang Y, Song F, Zhang W, et al. Analysis of immune-related signatures of lung adenocarcinoma identified two distinct subtypes: implications for immune checkpoint blockade therapy. Aging (Albany NY). 2020;12: 3312-3339. doi: 10.18632/aging.102814.
24. Herbst R.S, Giaccone G, Marinis F, Reinmuth N, Vergnenegre N, Barrios C.H, et al. Atezolizumab for first-line treatment of PD-L1–Selected patients with NSCLC. N Engl J Med. 2020;383:1328–1339. Doi:10.1056/NEJMoa1917346.
25. Gandhi L, Delvys R.A, Gadgeel S, Esteban E, Felip E, Angelis F.D, et al. Pembrolizumab plus chemotherapy in metastatic non-small-Cell lung Cancer. N Engl J Med. 2020;378:2078–2092. Doi:10.1056/NEJMoa1801005.

26. Hellmann M.D, Luis P.A, Caro R.B, Zurawski B, Kim S.W, Carcereny C.E, et al. Nivolumab plus ipilimumab in advanced non-small-Cell lung Cancer. N Engl J Med. 2019;381:2020–2031. Doi:10.1056/NEJMoa1910231.
27. Passaro A, Attili L, Morganti S, Signore E.D, Gianoncelli L, Spitaleri G, et al. Clinical features affecting survival in metastatic NSCLC treated with immunotherapy: a critical review of published data. Cancer Treat Rev. 2020;89:102085. doi: 10.1016/j.ctrv.2020.102085.
28. Carbone D.P, Reck M, Paz-Ares L, Creelan B, Horn L, Steins M, et al. First-line nivolumab in stage IV or recurrent non-small-cell lung cancer. N Engl J Med. 2017;376:2415-2426. doi: 10.1056/NEJMoa1613493.
29. Fehrenbacher L, Spira A, Ballinger M, Creelan B, Horn L, Steins M, et al. Atezolizumab versus docetaxel for patients with previously treated non-small-cell lung cancer (POPLAR): a multicentre. open-label. phase 2 randomised controlled trial. Lancet. 2016;10030:1837-1846. doi: 10.1016/ S0140-6736(16)00587-0.
30. Rittmeyer A, Barlesi F, Waterkamp D, Park K, Ciardiello F, Pawel J.V, et al. Atezolizumab versus docetaxel in patients with previously treated non-small-cell lung cancer (OAK): a phase 3. open-label. multicentre randomised controlled trial. Lancet. 2017;389:255-265. doi: 10.1016/S0140-6736(16)32517-X.
31. Powell S.F, Abreu D.R, Langer C.J, Tafreshi A, Ares L.P, Koppet HG, et al. 1483PD - Pembrolizumab (pembro) plus platinum-based chemotherapy (chemo) in NSCLC with brain metastases: Pooled analysis of KEYNOTE-021. 189. and 407. Annals of Oncology. 2019;30:v606-v607. doi:10.1093/annonc/mdz260.005.
32. Overman M.J, Lonardi S, Wong K.Y.M, Lenz H.J, Gelsomino F, Aglietta M, et al. Durable clinical benefit with nivolumab plus ipilimumab in DNA mismatch repair-deficient/ microsatellite instability-high metastatic colorectal cancer. J Clin Oncol. 2018;36:773-779. doi: 10.1200/JCO.2017.76.9901.
33. Hanahan D, Weinberg RA. Hallmarks of cancer the next generation. Cell. 2011;144:646-674. doi: 10.1016/j.cell.2011.02.013.
34. Fridman W.H, Pages F, Sautes-Fridman C, Galon J, et al. The immune contexture in human tumours: impact on clinical outcome. Nat Rev Cancer. 2012;14:298-306. doi: 10.1038/nrc3245.
35. Bremnes R.M, Al-Shibli K, Donnem T, Sirera R, Samer A.S, Andersen S, et al. The role of tumor-infiltrating immune cells and chronic inflammation atthe tumor site on cancer development. progression and prognosis emphasis on non-small cell lung cancer. J Thorac Oncol. 2011;6:824-833. doi: 10.1097/JTO.0b013e3182037b76.
36. Salgaller M.L. The development of immunotherapies for non-small cell lung cancer. Expert Opin Biol Ther. 2002; 2: 265-278. doi: 10.1517/14712598.2.3.265.
37. Remark R, Becker C, Gomez J.E, Damotte D, Dieu-Nosjean M.C, Fridman C.S, et al. The non-small cell lung cancer immune contexture. A major determinant of tumor characteristics and patient outcome. Am J Respir Crit Care Med. 2015; 191:377-390. doi: 10.1164/rccm.201409-1671PP.

38. Giraldo N.A, Becht E, Remark R, Damotte D, Sautès-Fridman C, Fridman WH. The immune contexture of primary and metastatic human tumours . Curr Opin Immunol. 2014;27:8-15. doi: 10.1016/j.coi.2014.01.001.
39. Fridman W.H, Remark R, Goc J, Giraldo N.A, Becht E, Hammond S.A, et al. The immune microenvironment a major player in human cancers . Int Arch Allergy Immunol. 2014;164:13-26. doi: 10.1159/000362332.
40. Bremnes RM, Busund LT, Kilvaer TL, Andersen S, Richardsen E, Paulsen EE, et al. The role of tumor-infiltrating lymphocytes in development progression and prognosis of non-small cell lung cancer . J Thorac Oncol. 2016;11:789-800. doi: 10.1016/j.jtho.2016.01.015.
41. Shah N, Sukumar S. The Hox genes and their roles in oncogenesis. Nat Rev Cancer. 2010;10:361-71. doi: 10.1038/nrc2826.
42. Zhang J, Yang M, Li D, Zhu S.Q, Zou J, Xu S.S, et al. Homeobox C8 is a transcriptional repressor of E-cadherin gene expression in non-small cell lung cancer. Int J Biochem Cell Biol. 2019;114:105557. doi: 10.1016/j.biocel.2019.06.005.
43. Liu H, Zhang M, Xu S, Zhang J, Zou J, Yang C, et al. HOXC8 promotes proliferation and migration through transcriptional up-regulation of TGFbeta1 in non-small cell lung cancer. Oncogenesis. 2018;7:1. doi: 10.1038/s41389-017-0016-4.
44. Yu M.J, Yu S.J, Zhou W, Yi B, Liu Y.H. HOXC6/8/10/13 predict poor prognosis and associate with immune infiltrations in glioblastoma. Int Immunopharmacol. 2021;101:108293. doi: 10.1016/j.intimp.2021.108293.
45. Lee TL, Shyu YC, Hsu PH, Chang CW, Wen SC, HsiaoWY, et al. JNK-mediated turnover and stabilization of the transcription factor p45/NF-E2 during differentiation of murine erythroleukemia cells. Proc Natl Acad Sci USA. 2010;107: 52-7. doi: 10.1073/pnas.0909153107.
46. Kapralova K, Lanikova L, Lorenzo F, Song Y.H, Horvathova M, Divoky V, et al. RUNX1 and NF-E2 upregulation is not specific for MPNs. but is seen in polycythemic disorders with augmented HIF signaling. Blood. 2014;123:391-4. doi: 10.1182/blood-2013-10-534222.
47. Wang X.S, Prensner J.R, Chen G, Qi Cao, Bo Han, Dhanasekaran S.M, et al. An integrative approach to reveal driver gene fusions from paired-end sequencing data in cancer. Nature Biotechnology. 2009;27. 1005-11. doi: 10.1038/nbt.1584
48. Dou R, Wang X, Zhang J. Prognostic Value and Immune Infiltration Analysis of Nuclear Factor Erythroid-2 Family Members in Ovarian Cancer. Biomed Res Int. 2022;8672258. doi: 10.1155/2022/8672258.
49. Klebanoff CA, Gattinoni L, Restifo NP. Sorting through subsets: which T-cell populations mediate highly effective adoptive immunotherapy? J Immunother. 2012;35 651-60. doi: 10.1097/CJI.0b013e31827806e6.
50. Klebanoff C.A, Gattinoni L, Parizi P.T, Kerstann K, Cardones A.R, Finkelstein S.E, et al. Central Memory self/tumor-reactive CD8 T cellsconfer superior antitumor immunity compared with effector memory T cells. Proc Natl Acad Sci USA; 2005;102:9571-6. doi: 10.1073/pnas.0503726102.

51. Lee M.H, Murphy G. Matrix metalloproteinases at a glance. J Cell Sci. 2004;117:4015-6. doi: 10.1242/jcs.01223.
52. Bloomston M, Zervos E.E, Rosemurgy A.S. Matrix metalloproteinases and their role in pancreatic cancer: a review of preclinical studies and clinical trials. Ann Surg Oncol. 2002;9:668-74. doi: 10.1007/BF02574483.
53. Stetler-Stevenson W.G. Progelatinase A activation during tumor cell invasion. Invasion Metastasis. 1994;14: 259-68.
54. Kleiner D.E, Stetler-Stevenson W.G. Matrix metallo- proteinases and metastasis. Cancer Chemother Phar-macol. 1999;43: S42-51.
55. Sternlicht M.D, Werb Z. How matrix metalloprotei- nases regulate cell behavior. Annu Rev Cell Dev Biol. 2001;17: 463-516.
56. Hofmann H.S, Hansen G, Richter G, Taege C, Simm A, Silber R.E, et al. Matrix Metalloproteinase-12 Expression Correlates with Local Recurrence and Metastatic Disease in Non–Small Cell Lung Cancer Patients. Clin Cancer Res. 2005; 11:1086-92.
57. Hung W.Y, Lee W.J, Cheng G.Z, Tsai C.H, Yang Y.C, Lai T.C, et al. Blocking MMP-12-modulated epithelial-mesenchymal transition by repurposing penfluridol restrains LUAD metastasis via uPA/uPAR/TGF-β/Akt pathway. Cellular Oncology. 2021;44:1087-1103. doi: 10.1007/s13402-021-00620-1.
58. Quan X, Liu X, Ye D.M, Ding X.L, Su X.L. Forsythoside A Alleviates High Glucose-Induced Oxidative Stress and Inflammation in Podocytes by Inactivating MAPK Signaling via MMP12 Inhibition. Diabetes Metab Syndr Obes. 2021; 28:14. 1885-1895. doi: 10.2147/DMSO.S305092.
59. Kwon C.H, Moon H.J, Park H.J, Ding X.L, Su X.L. S100A8 and S100A9 promotes invasion and migration through p38 mitogen-activated protein kinase-dependent NF-κB activation in gastric cancer cells. Molecules & Cells. 2013;3:226-234. doi: 10.1007/s10059-013-2269-x.
60. Su W, Fan H, Chen M, Wang J.L, Brand D, He X.S, et al. Induced CD4+ forkhead box protein–positive T cells inhibit mast cell function and established contact hypersensitivity through TGF-β1. Journal of Allergy & Clinical Immunology. 2012; 130:444-452. doi: 10.1016/j.jaci.2012.05.011.
61. Kim B.S, Clinton J, .Wang Q, Chang S.H. Targeting ST2 expressing activated regulatory T cells in Kras-mutant lung cancer. Oncoimmunology 2019;9:1682380. doi: 10.1080/2162402X.2019.1682380.

Figures

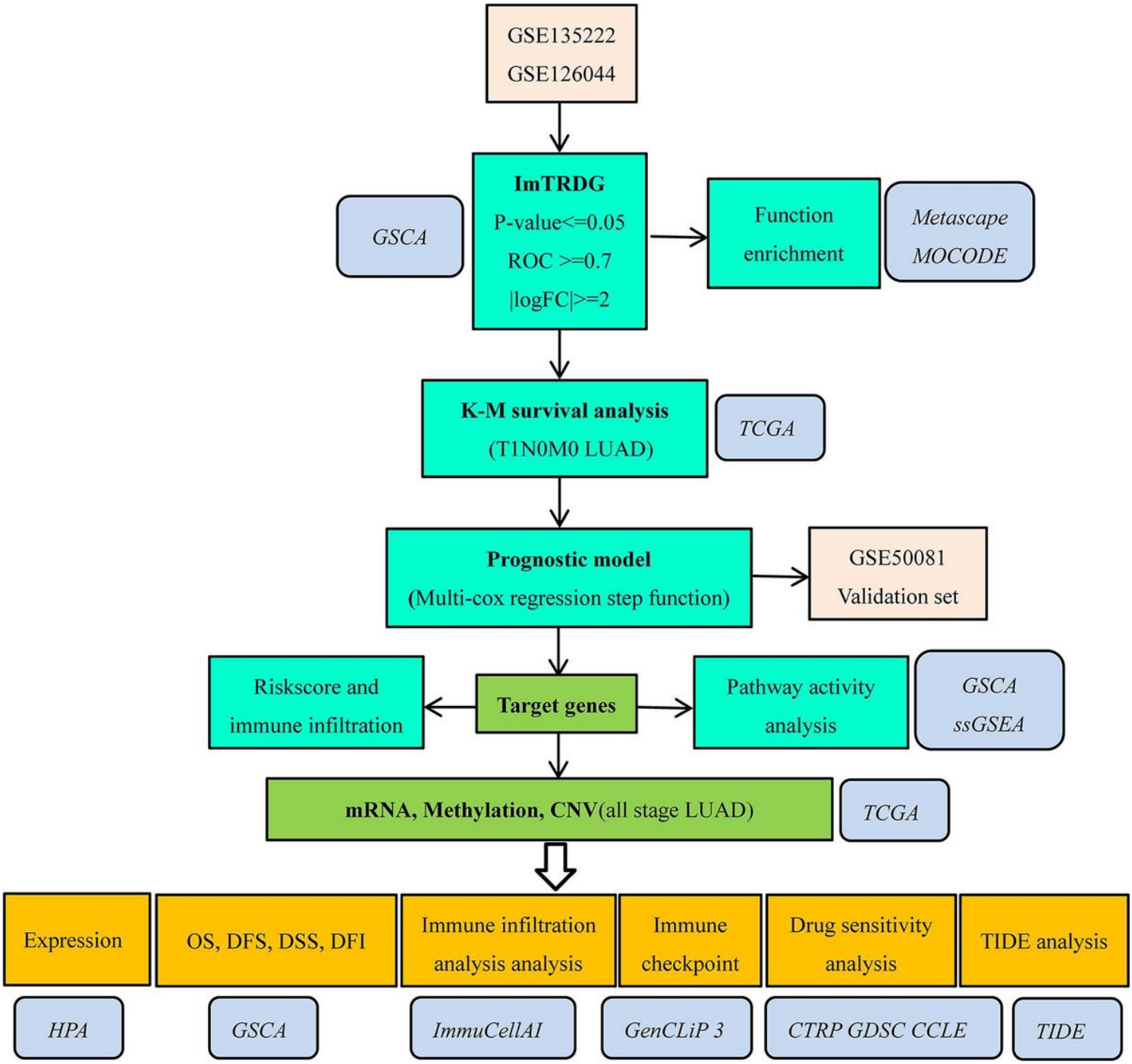

Figure 1

The overall process of research contents.

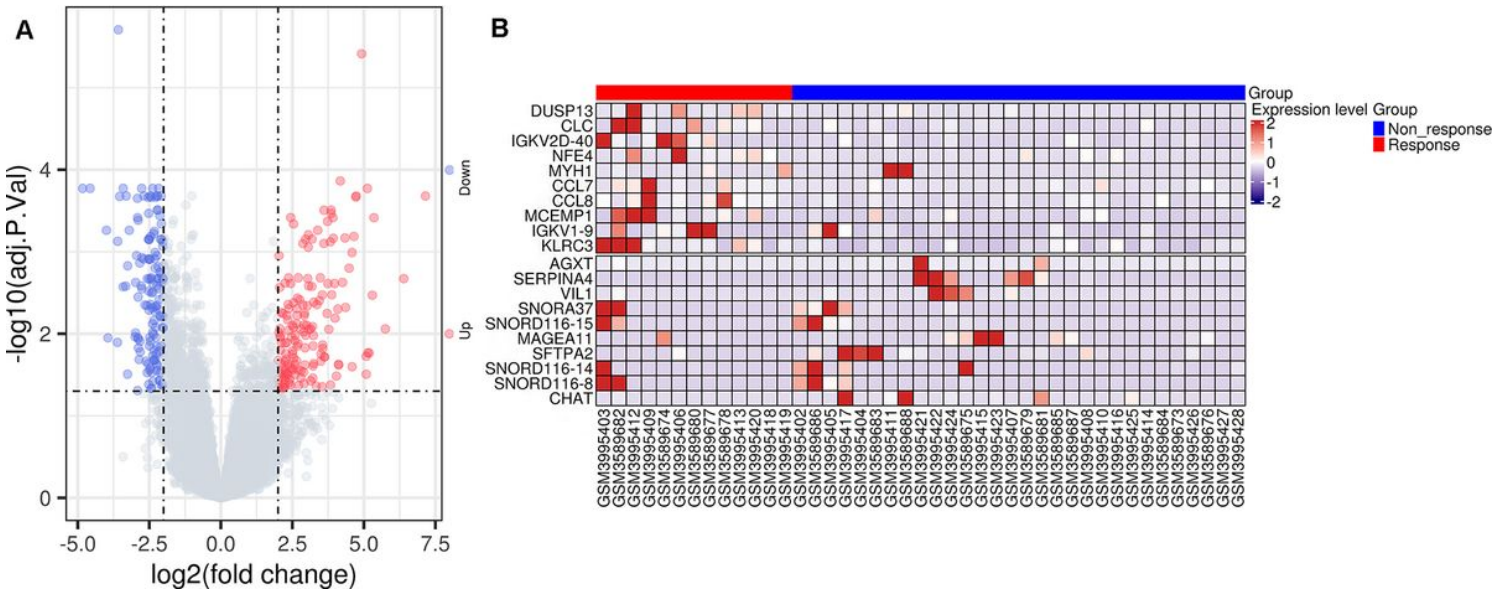

Figure 2

Visual presentation of imTRDG A. Volcano plot of ImTRDG for p<0.05, |LogFC>2; B. Heatmap of the top 20 ImTRDGs.

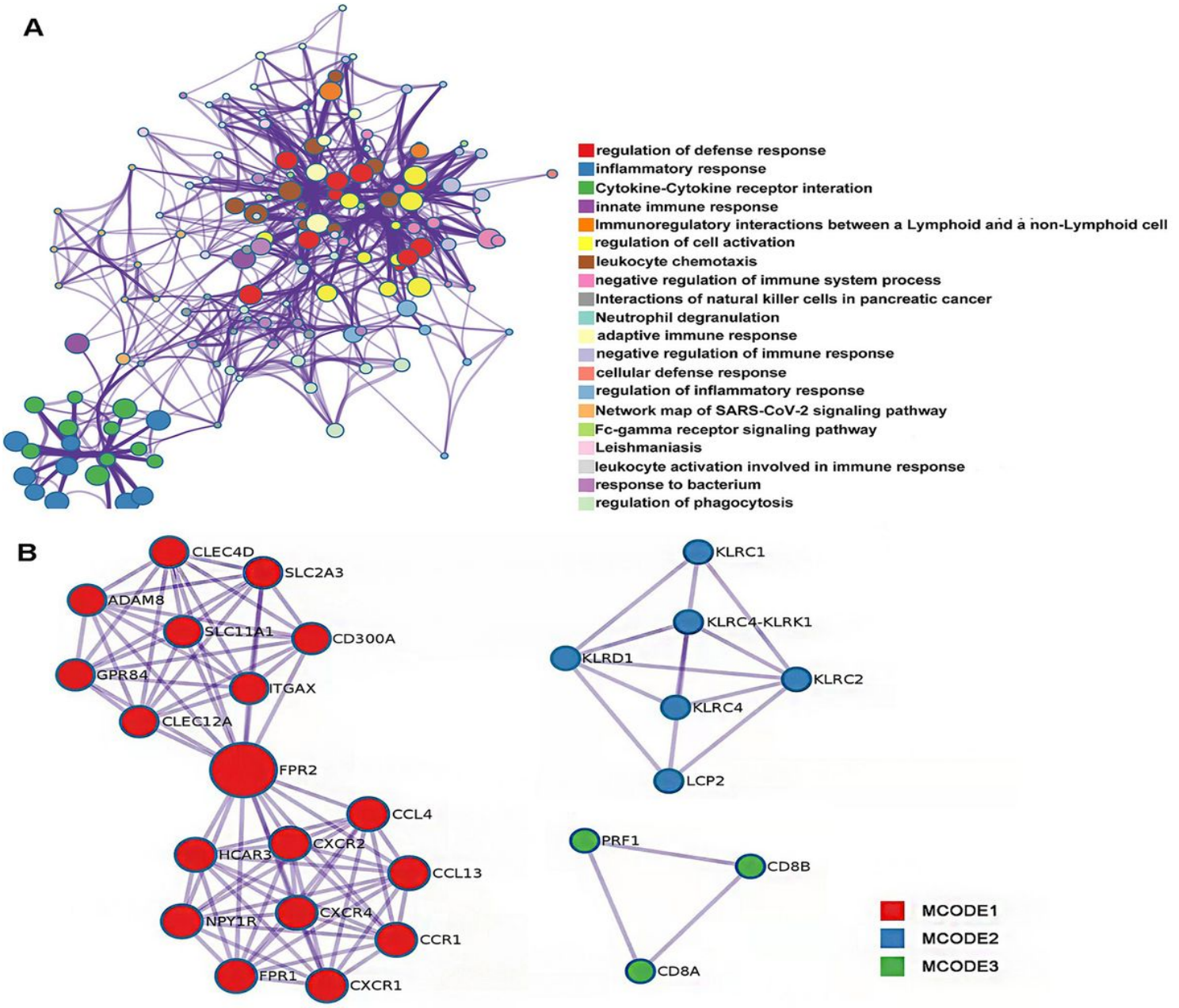

Figure 3

Visual presentation of imTRDG functional enrichment results. A. Enriched Ontology Clusters Colored by Cluster type; B. PPI MCODE components network.

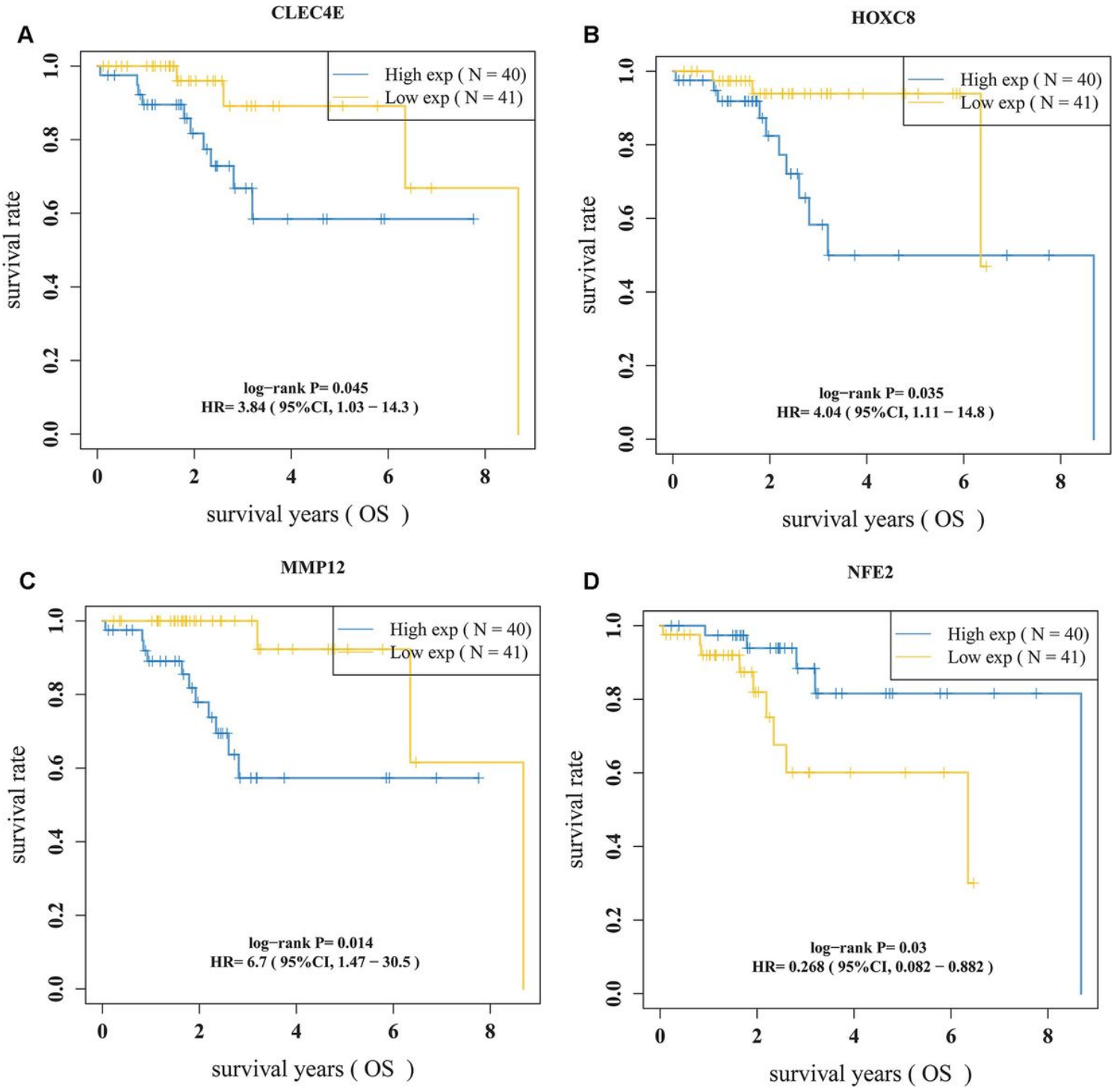

Figure 4

Univariate K-M survival curves of target genes. A. K-M survival curves of CLEC4E in T1N0M0 LUAD; B. K-M survival curves of HOXC8 in T1N0M0 LUAD; C. K-M survival curves of MMP12 in T1N0M0 LUAD; D. K-M survival curves of NFE2 in T1N0M0 LUAD.

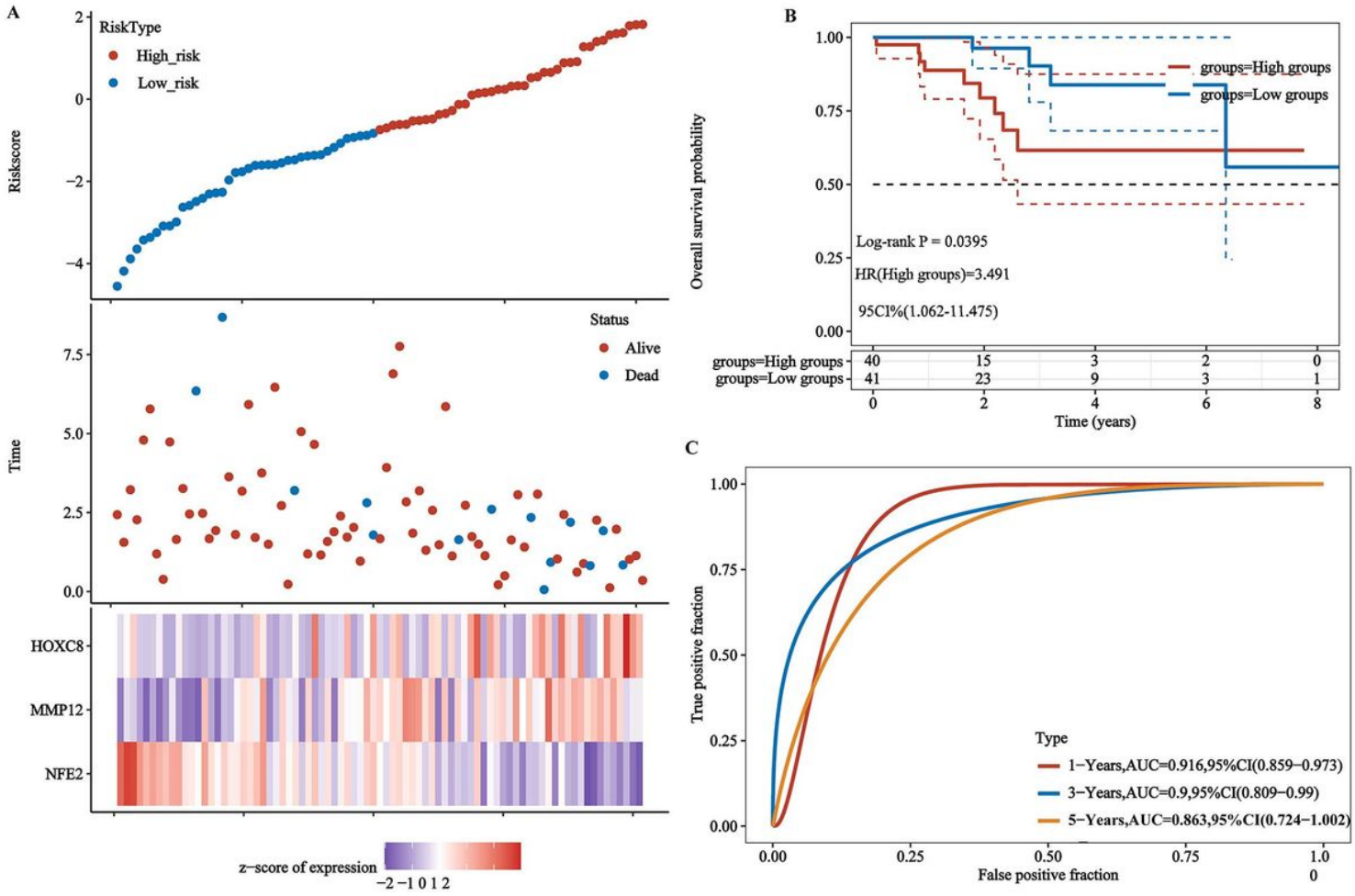

Figure 5

Multivariate K-M survival curves and ROC curves of risk model. A. The survival time, survival status scatter plot and target gene distribution heat map corresponding to riskscore of different samples. The abscissa and ordinate represent genes, in which different colors represent the correlation coefficient, and the darker the color, the stronger the correlation; B. Distribution of KM survival curves by prognostic risk model in T1N0M0 LUAD; C. ROC curves and AUC values at different times in T1N0M0 LUAD risk model.

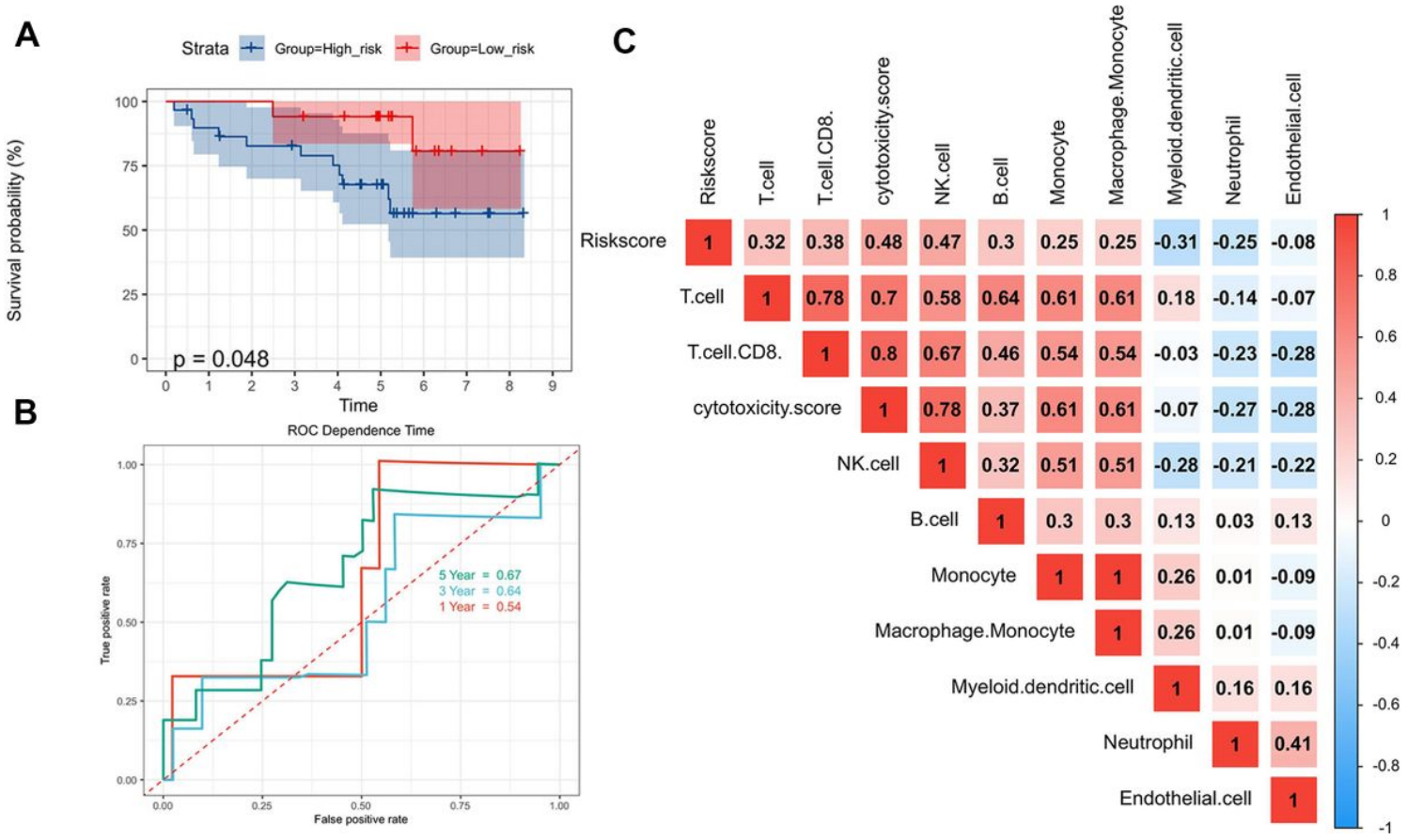

Figure 6

Multivariate K-M survival curves, ROC curves and immune infiltration landscapes of model genes in GSE50081 validation set. A. Distribution of KM survival curves by prognostic risk model in GSE50081 T1N0M0 LUAD; B. ROC curves and AUC values at different times in GSE50081 T1N0M0 LUAD risk model; C. The immune infiltration landscapes corresponding to riskscore in GSE50081 validation set.

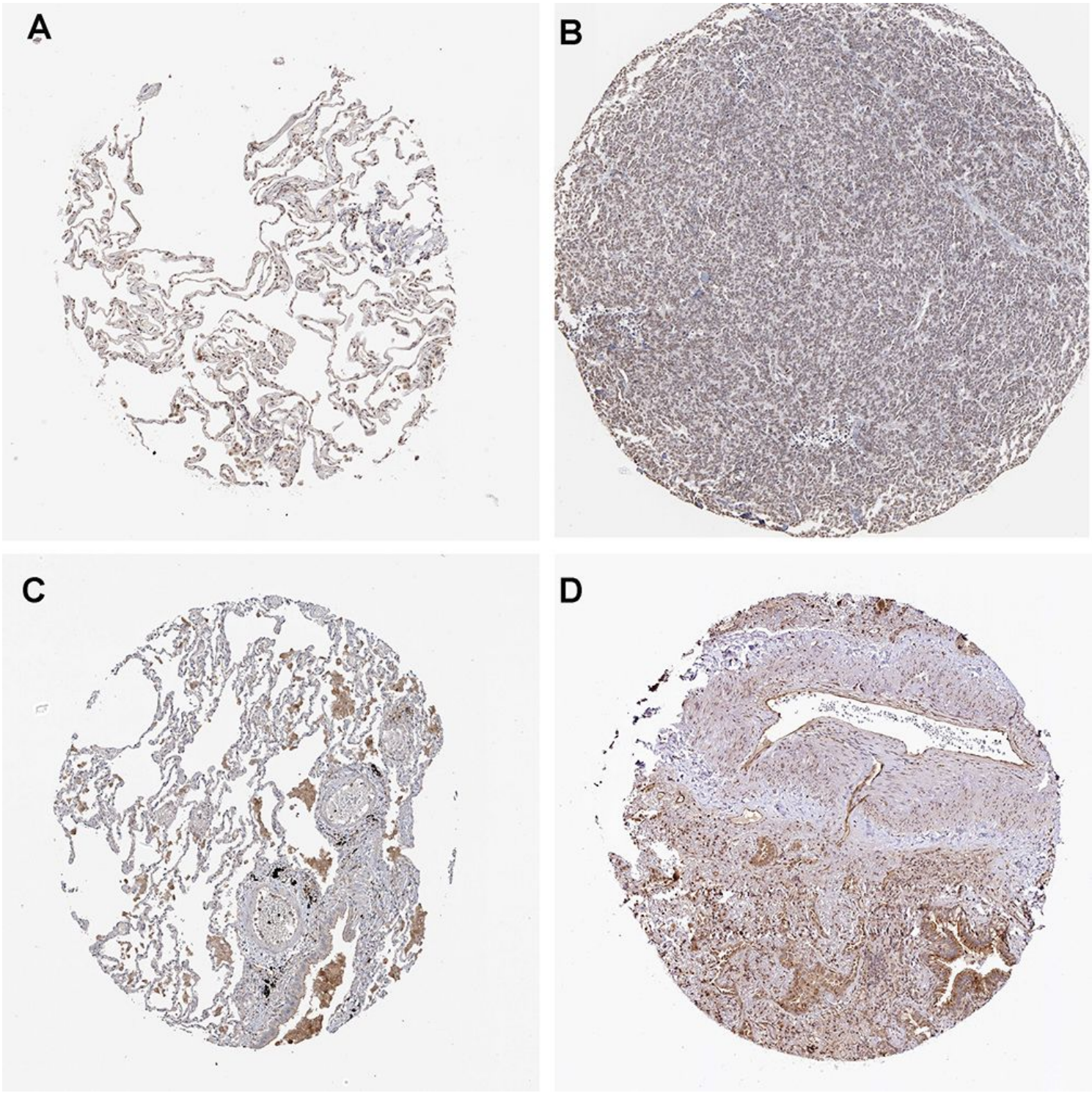

Figure 7

Immunohistochemical results of target genes proteins in normal lung and tumor tissues from the HPA database. A. Immunohistochemical profile of NFE2 in normal lung; B. Immunohistochemical profile of NFE2 in LUAD; C. Immunohistochemical profile of HOXC8 in normal lung; D. Immunohistochemical profile of HOXC8 in LUAD.

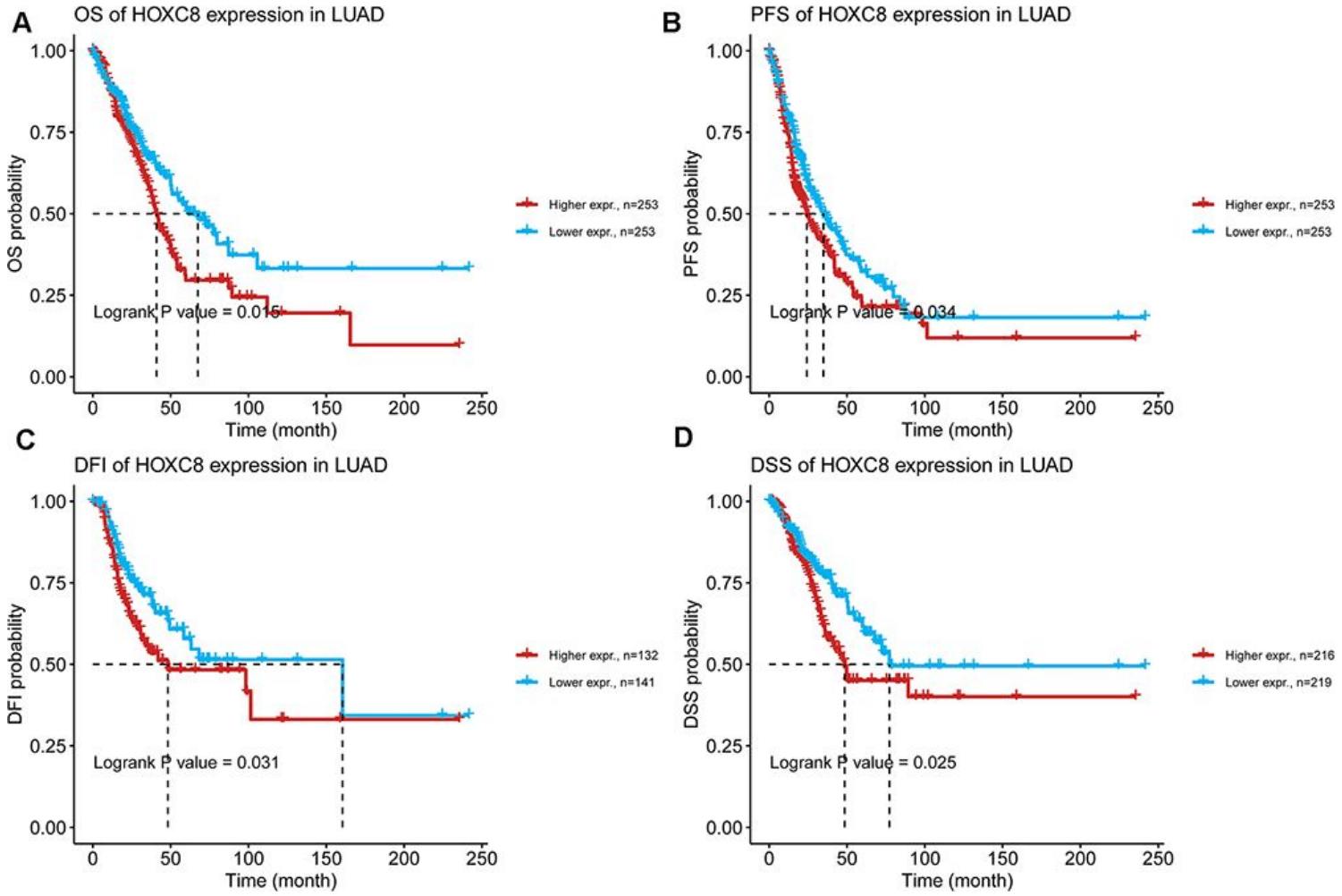

Figure 8

The multi-type K-M survival curves of target genes in LUAD. A. OS K-M curves of HOXC8 in all stages of LUAD; B. PFS K-M curves of HOXC8 in all stages of LUAD; C. DFI K-M curves of HOXC8 in all stages of LUAD. D. DSS K-M curves of HOXC8 in all stages of LUAD. OS. overall survival, PFS. progression free survival, DSS. disease specific survival, DFI. Disease Free Interval.

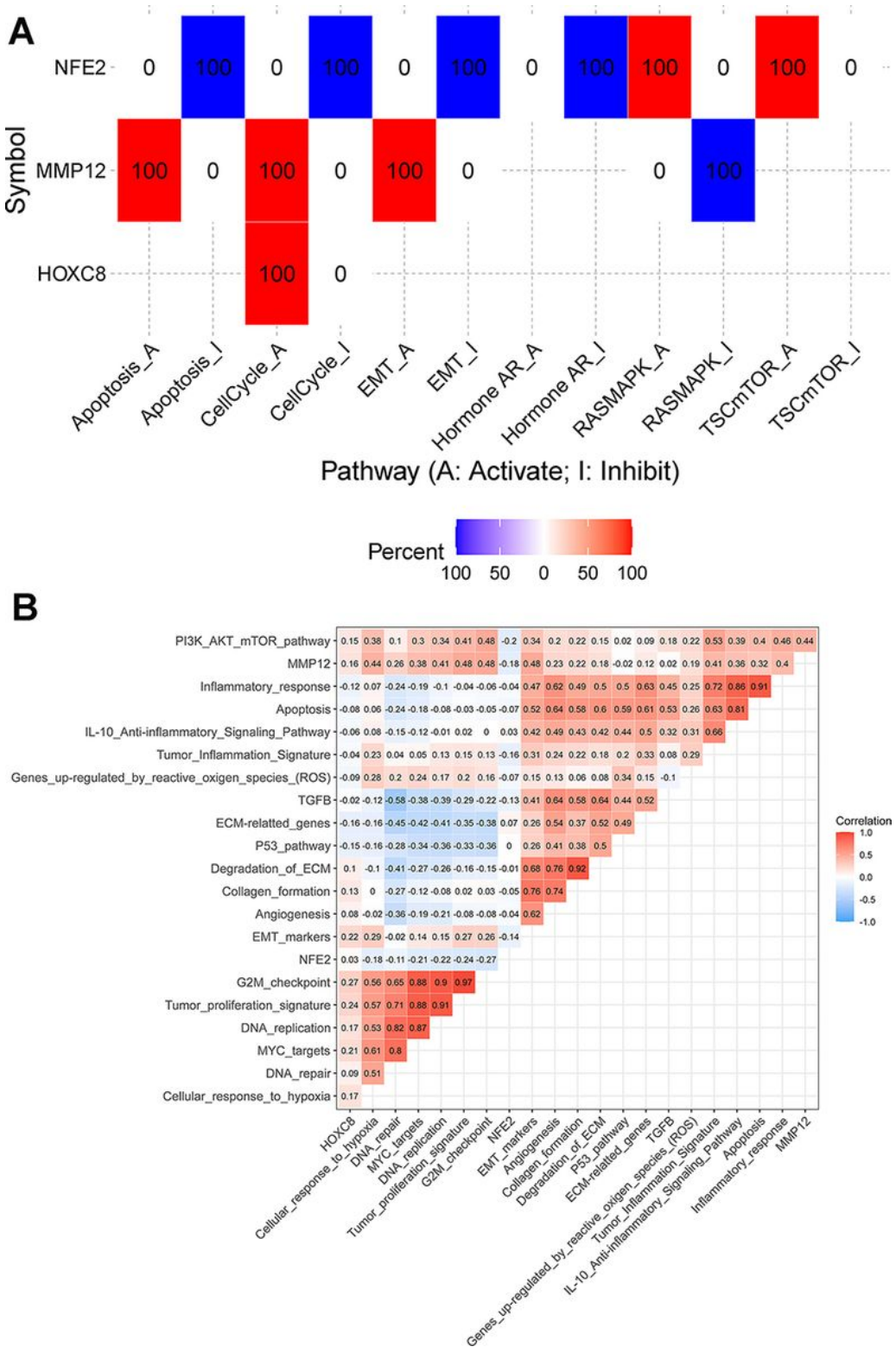

Figure 9

Heatmap of the correlation between target genes and immune infiltration. A. Correlation heatmap of expression and pathway activity from GSCA. Percentage of cancer in which a gene have effect (FDR <=0.05) on the pathway in LUAD, the number in each cell indicates the percentage; B. Correlation heatmap of gene expression and pathway activity from ssGSEA.

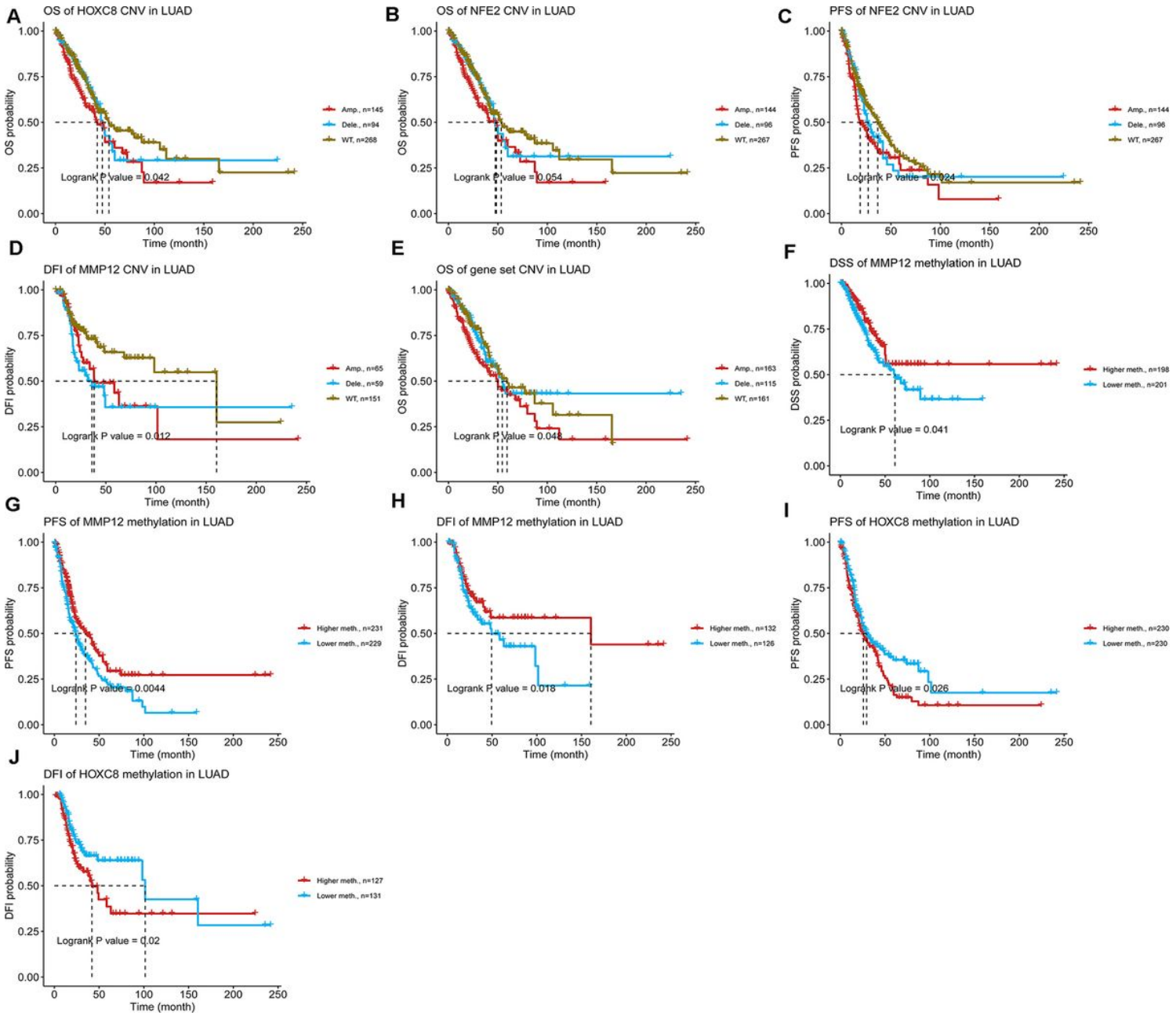

Figure 10

K-M survival prognosis curve of Copy Number Variation (CNV) and methylation of target genes. A.OS of HOXC8 CNV in LUAD; B. OS of NFE2 CNV in LUAD; C. PFS of NFE2 CNV in LUAD; D. DFI of MMP12 CNV in LUAD; E. OS of gene set CNV in LUAD; F. DSS of MMP12 methylation in LUAD; G. PFS of MMP12 methylation in LUAD; H. DFI of MMP12 methylation in LUAD; I. PFS of HOXC8 methylation in LUAD; J. DFI of HOXC8 methylation in LUAD. OS. overall survival, PFS. progression free survival, DSS. disease specific survival, DFI. Disease Free Interval.

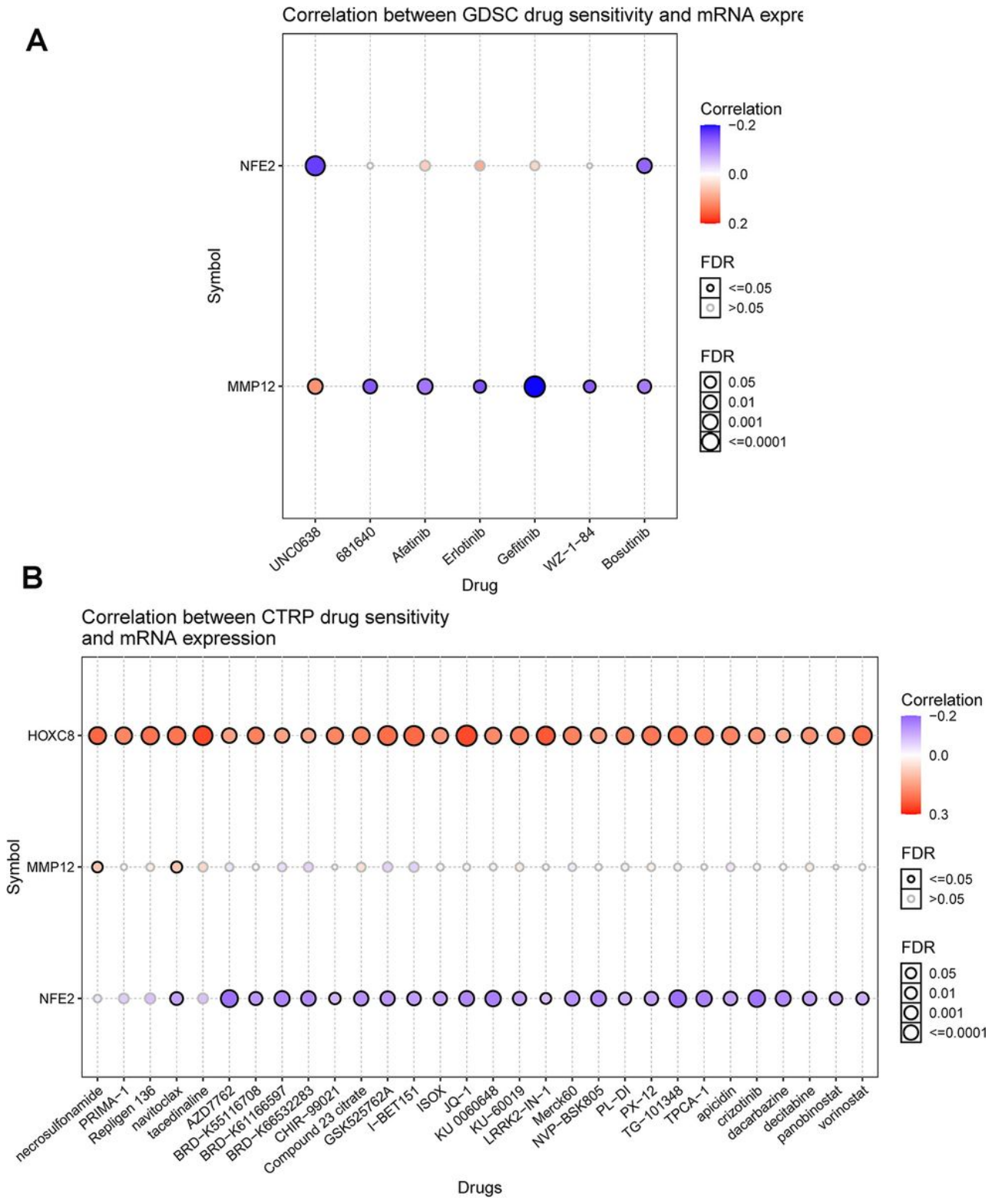

Figure 11

Distribution of IC50 of different drugs between high and low expression groups of target genes. A. Distribution of IC50 of different drugs between high and low expression groups of target genes from GDSC; B. Distribution of IC50 of different drugs between high and low expression groups of target genes from CTRP.

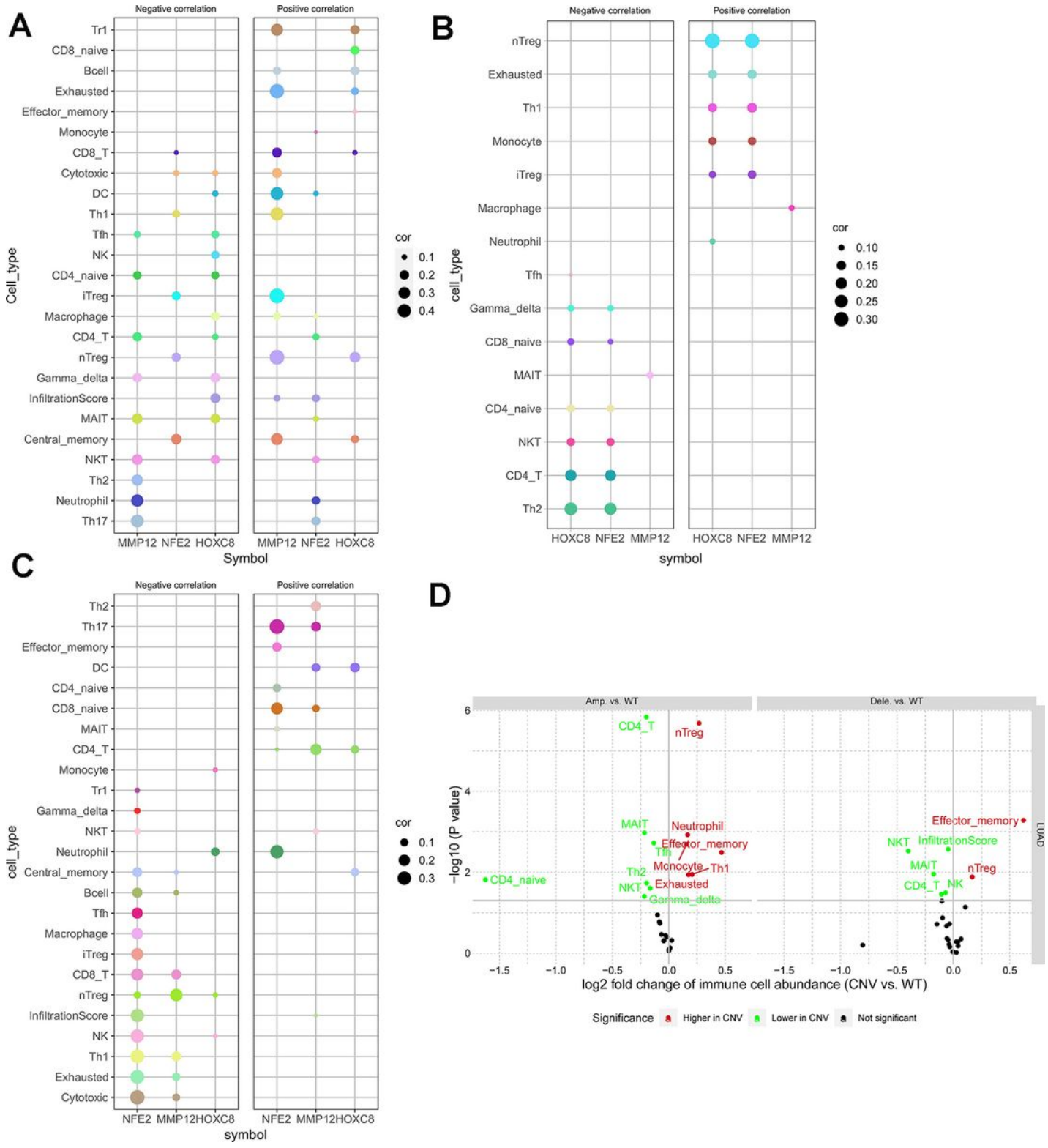

Figure 12

Bubble plot of the correlation between target genes and immune infiltration. A. The correlation between target gene mRNA expression and immune infiltration; B. The correlation between methylation level of target gene and immune infiltration; C. The correlation between CNV status of target gene and immune infiltration; D. The correlation of gene set CNV correlation with immune infiltration. The correlation P values in the above figures are all less than 0.05, which is statistically significant.

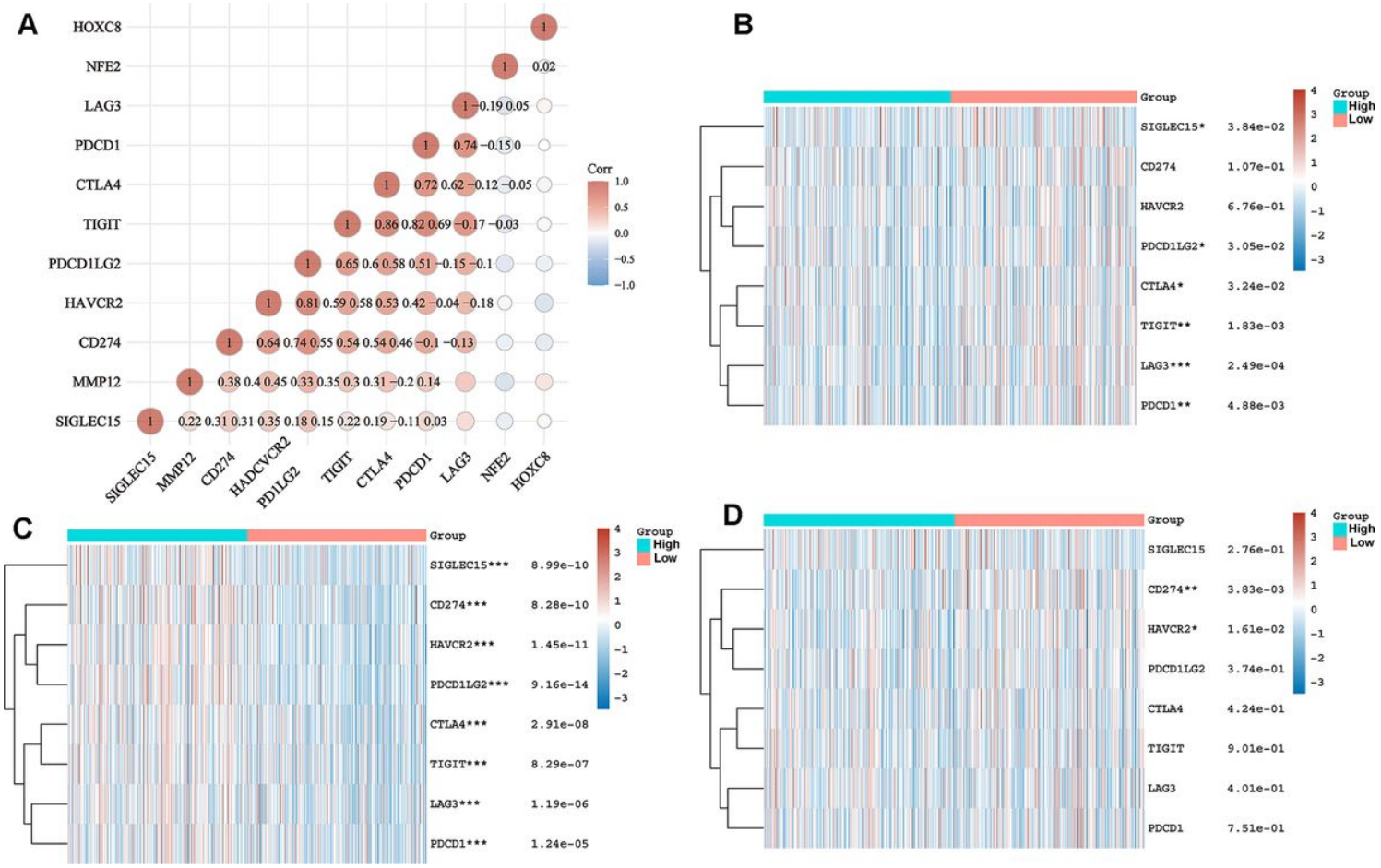

Figure 13

Heatmap of correlations between target genes and immune check-related genes. A. Heatmap of linear correlations between target genes and immune check-related genes; B. Heat map of differential expression of immune checkpoint-related genes between high and low HOXC8 expression groups; C. Heat map of differential expression of immune checkpoint-related genes between high and low NFE2 expression groups; D. Heat map of differential expression of immune checkpoint-related genes between high and low MMP12 expression groups. *p < 0.05, **p < 0.01,***p < 0.001, asterisks (*) stand for significance levels.

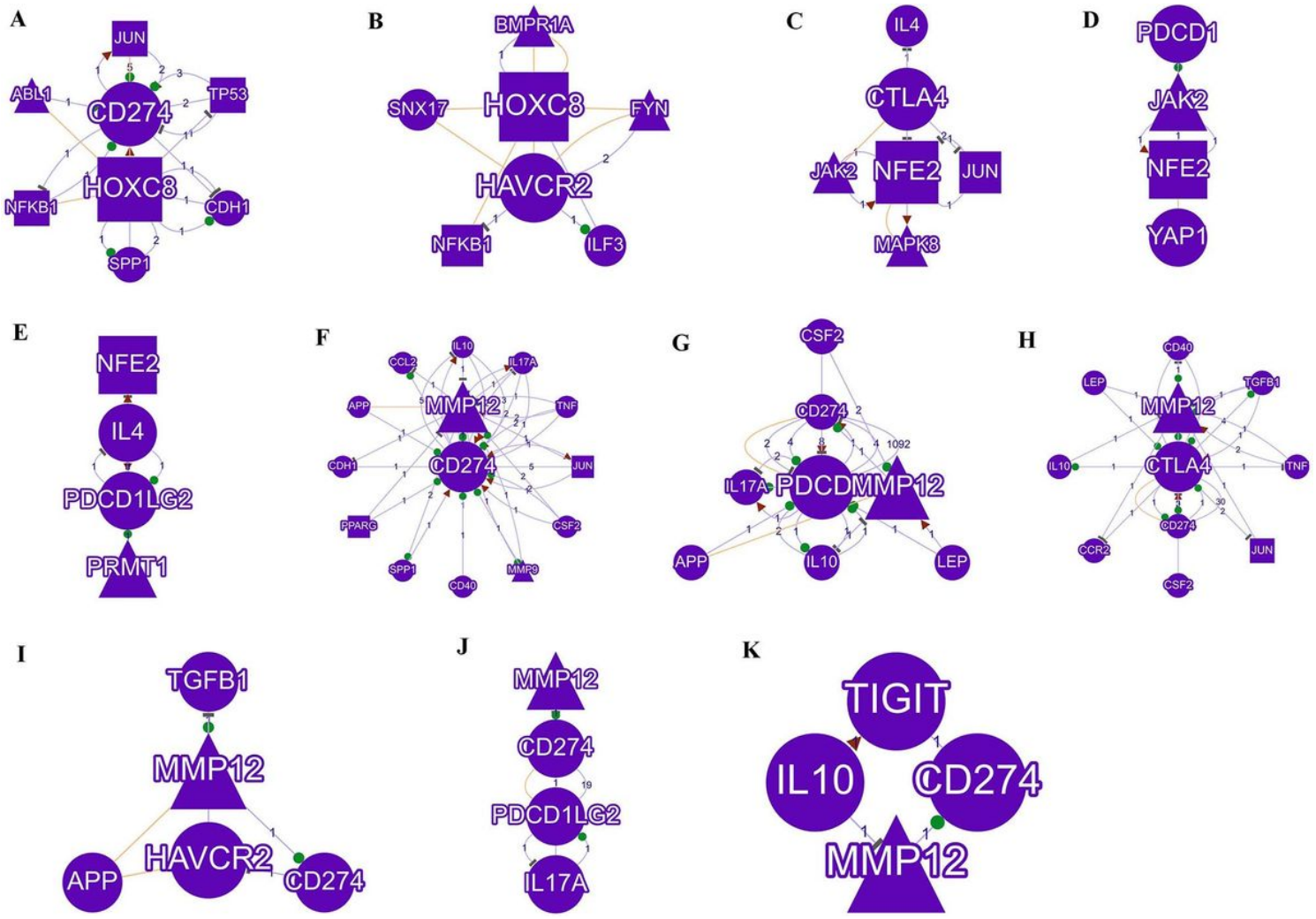

Figure 14

Regulatory networks of target genes and immune checkpoint-related genes. A. Cross-talk between HOXC8 and CD274; B. Cross-talk between HOXC8 and HAVCR2; C. Cross-talk between NFE2 and CTLA4; D. Cross-talk between NFE2 and PDCD1; E. Cross-talk between NFE2 and PDCD1LG2; F. Cross-talk between MMP12 and CD274; G. Cross-talk between MMP12 and PDCD1; H. Cross-talk between MMP12 and CTLA4; I. Cross-talk between MMP12 and HAVCR2; J. Cross-talk between MMP12 and PDCD1LG2; K. Cross-talk between MMP12 and TIGIT. Circles represent genes, triangles represent enzymes, squares represent transcription factors, lines represent interactions, and the numbers on the lines represent the number of studies of interactions between genes that were experimentally validated or data-mined.

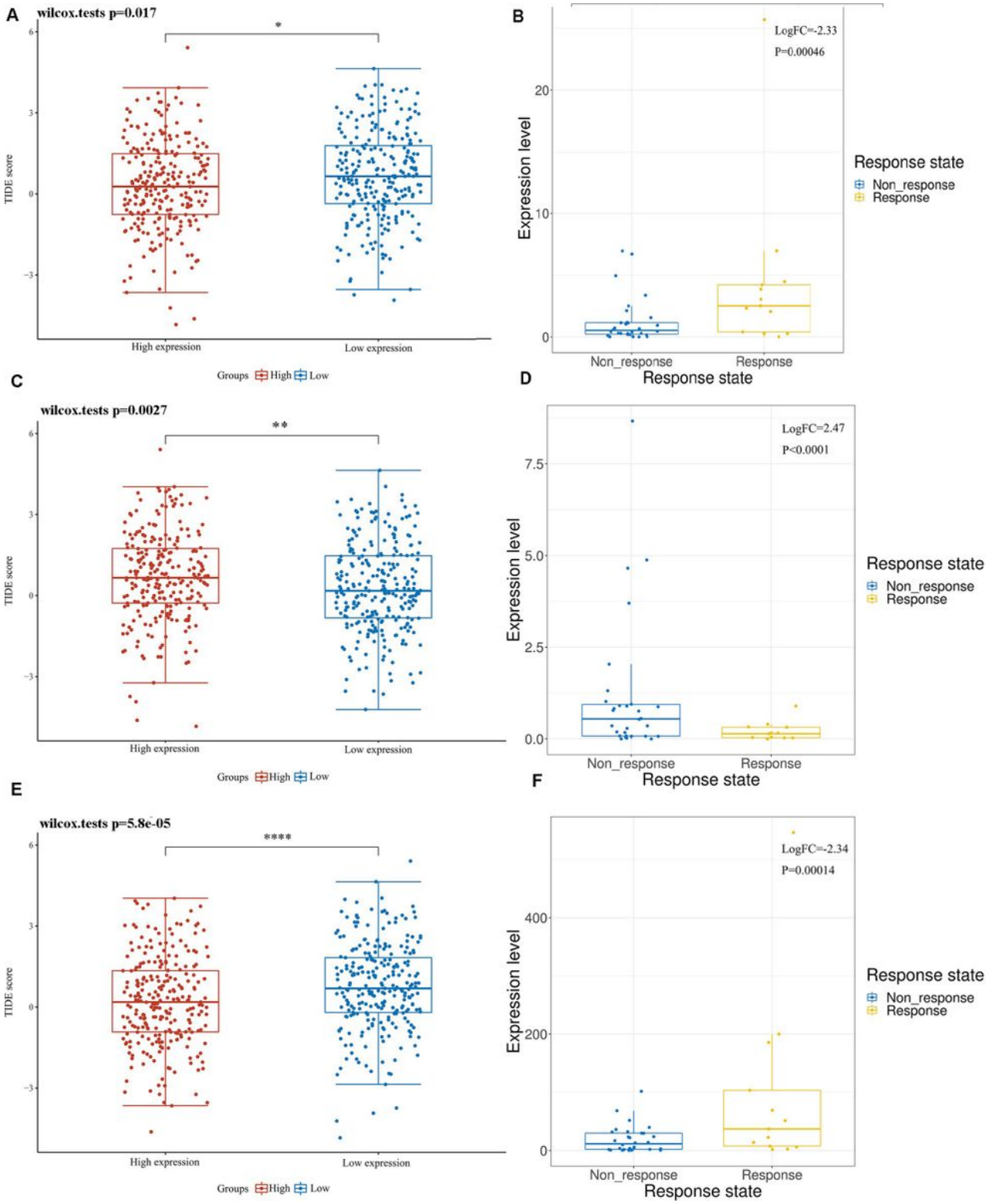

Figure 15

Prediction of the effect of immunotherapy under different expression of target genes. A. Distribution of immune responses and immune scores in NFE2 high and low groups; B. Distribution of NFE2 between immunotherapy responders and non-responders in the CTR-DB database; C. Distribution of immune responses and immune scores in HOXC8 high and low groups; D. Distribution of HOXC8 between immunotherapy responders and non-responders in the CTR-DB database; E. Distribution of immune

responses and immune scores in MMP12 high and low groups. F. Distribution of MMP12 between immunotherapy responders and non-responders in the CTR-DB database. High TIDE score, poor response to immune checkpoint blockade (ICB), and short survival after ICB. *p < 0.05, **p < 0.01,***p < 0.001, asterisks (*) stand for significance levels.

Supplementary Files

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

- SupplementaryFigure1.jpg
- SupplementaryFigure3.jpg
- SupplementaryFigure4.jpg
- SupplementaryFigure2.jpg
- Supplementarytable10ImmuneAndMethyTable.xlsx
- Supplementarytable11GdscIC50AndExprTable.xlsx
- Supplementarytable12CtrpDrugIC50AndExpTable.xlsx
- Supplementarytable13MMP12ccletargetresponsediff.xlsx
- Supplementarytable14HOXC8ccletargetresponsediff.xlsx
- Supplementarytable1ImTRDG.csv
- Supplementarytable2metascapeenrichmentanalysis.csv
- Supplementarytable3PPIMCODEcomponents.xlsx
- Supplementarytable4GSE50081data.xlsx
- Supplementarytable5Riskscoreimmuneinfiltration.xlsx
- Supplementarytable6GeneexpressionandpathwayactivityresultfromGSCA.xlsx
- Supplementarytable7pathwayssGSEAScore.xlsx
- Supplementarytable8ImmuneAndExprTable.xlsx
- Supplementarytable9ImmuneAndCnvTable.xlsx
- Supplementaryfigurelegends.docx