

Distinct Expression Pattern and Prognostic Values of E2F Family Genes in Lung Adenocarcinoma and Squamous Cell Carcinoma

Shuo Wang (✉ rrr71@126.com)

THE 2nd Hospital of Tianjin Medical University

Jun Zhang

The 2nd Hospital of Tianjin Medical University

Fan-Jie Meng

The 2nd Hospital of Tianjin Medical University

Yi-Jie Yan

The 2nd hospital of tianjin medical university

Bin Wang

The 2nd hospital of tianjin medical university

Zhi-Yu Guan

Tianjin Medical University

Research

Keywords: Distinct Expression Pattern, Prognostic Values, E2F Family Genes , Lung Adenocarcinoma , Squamous Cell Carcinoma

Posted Date: November 11th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-1015669/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

The E2F transcription factors family included E2F1-8 playing the crucial roles in the origination and progression of various kinds of human cancers. However, a comprehensive analysis regarding the gene mRNA expression pattern at bulk and single-cell resolution, the prognostic values and the association to immune cell infiltration of E2F family genes in lung adenocarcinoma (LUAD) and squamous cell carcinoma (LUSC) remains to be performed. The bulk and single-cell mRNA expression levels of E2F family members in LUAD and LUSC were determined using the cBioPortal, Gepia, TISCH and Oncomine online databases. Based on the differentially expressed genes, GO enrichment and GSEA items were investigated. The Kaplan-Meier plotter server was used to evaluate the prognostic significances of the E2F family genes in patients with LUAD and LUSC based on the related clinicopathological features. The correlation between E2F family gene expression and immune cell infiltration was explored using the Timer database. Our bulk and single-cell RNA analyses revealed that E2F1, E2F2 and E2F8 might promote the tumor growth and aggressiveness of LUAD and LUSC indicating the poor prognosis prediction. E2F4 and E2F6 might be considered as the potential outcome markers for the improved treatment of LUAD and LUSC, respectively. Upregulation of E2F1 and E2F8 indicated the poor prognosis of LUAD patients, whereas only E2F2 overexpression was associated with shorter overall survival of LUSC patients. There existed a positive relationship between E2F8 expression level and infiltration level of macrophages and DCs, and significantly positive correlation between infiltration level of CD8+ T, CD4+ T cells and E2F1/2/7 expression in LUAD and LUSC. In conclusion, our findings suggested that increased expression of E2F1/E2F8 or E2F2 may serve as promising prognostic biomarkers for patients with LUAD or LUSC, respectively.

Introduction

Lung cancer originates from the cells that line the air passages as the leading cause of cancer-associated mortality around the worldwide [1]. There exist two main types: small cell lung cancer and non-small cell lung cancer (NSCLC). NSCLC included lung adenocarcinoma (LUAD) and lung squamous cell carcinoma (LUSC), accounting for about 80% of the entire lung cancers [2]. The tumor cells can spread to nearby tissues or other body systems beyond the lung, which leads to a five-year survival rate of <15% [3]. The 85% of lung cancer patients get sick from long-term tobacco smoking, and additionally about 15% of cases occur in the people never the smoked ones [4, 5]. During disease progression, the combination of genetic factors produces the crucial effects on the origination and recurrences [6].

Although several therapeutic approaches for NSCLC patients have made much advanced progresses, including the operation, chemotherapy, radiation and molecular targeted therapy [7], it is an unavoidable reality that the percentage of patients is still increasing [8]. Treatment benefits and long-term prognosis depend mainly on the molecular subtypes of cancer, tumor stages (degree of spreading) and the overall health [9]. Unfortunately, almost 80% of NSCLC patients at the first diagnosis have progressed into the advanced stages [10]. Thus, an urgent requirement is needed to deepen the understanding of the

mechanisms of the onset and progression of NSCLC and identify the potential prognostic genes. The construction of biomarkers in the diagnosis and treatment may improve the therapeutic strategies for NSCLC patients.

E2F family genes have been reported to function as the transcription factors in higher eukaryotes [11]. In the contemporary era, the onset of cancer can be explained by the development concept and the tumor microenvironment has been identified to be similar to the development microenvironment [12]. E2F family members play the critical roles in regulating the cell cycle during the G1/S transition by bind to the target promoter sequence [13], such as the TTTCCCGC binding site [14]. Amongst them, some have been identified as the activators, such as the E2F1 and E2F2, while some others were considered as the suppressors, such as E2F4 and E2F6 [15]. The vital functions of E2F family genes have been studied in various cancers, including the breast cancer, bladder cancer and lung cancer [16–18]. Previously, E2F1 has been reported to function in downregulating the expression of tumor suppressor gene ARHI in breast cancer via epigenetic regulation [19]. E2F3 has been reported to contribute to maintaining the stemness of bladder cancer stem-like cells by chromatin deregulation through enhancers or super-enhancers [20]. By interacting with the microRNA networks, E2F family genes also promoted the pediatric brain tumor progression [21]. Therefore, with the various mechanisms of regulating the downstream factors, such as the histone proteins, targeted genes and microRNAs, E2F family members extensively participate in cancer progression.

The study model of multi-omics has been applied in the lung cancer [22], which give us a more comprehensive understanding of this devastating disorder at the big data resolution. However, the underlying mechanism of E2F family genes and the distinct functions have yet to be fully illustrated in LUAD and LUSC, respectively. In the current study, based on the analyses of gene expressions or copy numbers variations by using the online available databases, the expression and mutations of E2F family genes in patients with LUAD and LUSC were analyzed in detail to identify the potential functions and prognostic values. The association with immune cell infiltration was focused on to reveal the impact of E2F family members on the tumor microenvironment. Furthermore, single-cell RNA sequence (scRNA-seq) data was included with the hope of specifying the expression and functions at the malignant and non-malignant clusters.

Materials And Methods

Genomic alteration of E2F family in the cBioPortal database

To investigate the genomic alterations of E2F family genes and the correlation between copy number variation (CNV) and mRNA expression, the cBioPortal database (<https://www.cbioportal.org>) was evacuated. Genomic alteration types and alteration frequencies were analyzed in various tumors including LUAD and LUSC. The genomic alterations of E2F family, such as CNV, deep deletion, missense mutation, truncating mutation with unknown significance, were provided by the the cBioPortal Oncoprint

tool. The correlation between the genetic alterations and the mRNA expression levels of E2F family was analyzed in LUAC and LUSC.

Gene expression analysis in GEPIA

The GEPIA (Gene Expression Profiling Interactive Analysis, <http://gepia.cancer-pku.cn>) database can provide the differential gene expression between normal and cancer tissues, which constructs the interactive web server for exploring the RNA sequencing expression data from 9736 tumors and 8587 normal tissues based on the TCGA and the GTEx (Genotype-Tissue Expression) programs [23]. The tumor and normal tissue datasets of various cancers including LUAD and LUSC were used for analyzing the expression of E2F family genes. Gene expression correlation analysis was performed in the LUAD and LUSC datasets from TCGA data. E2F family genes were used for the x-axis, and other genes of interest including the Ki-67 and EGFR are represented on the y-axis, of which the Spearman's correlation analysis was chosen to determine the correlation coefficient.

The immune infiltration analysis in TIMER database

TIMER database (<http://timer.cistrome.org>) has become a comprehensive online server for systematic analysis of immune infiltration among the diverse cancer types [24]. The TIMER database totally includes 10,897 samples across 32 cancer types from TCGA data to estimate the abundance of immune infiltrations. The association between immune infiltration and clinical outcome was evaluated by the Kaplan-Meier evaluation. In the TIMER database, a deconvolution was used to infer the abundance of tumor-infiltrating immune cells (TIICs) from gene expression profiles according to the previously published statistical methods. We analyzed the correlation of E2F family gene expression with the tumor purity and the abundance of immune infiltrates, including B cells, CD4⁺ T cells, CD8⁺ T cells, neutrophils, tumor-associated macrophages, and dendritic cells. In addition, correlations between prognosis and tumor-infiltrating immune cells were explored via log-rank test. E2F family was used for the x-axis with gene symbols, and the corresponding infiltrating immune cells are represented on the y-axis. The gene expression level was displayed with log₂ RSEM.

The survival analysis based on the Kaplan-Meier Plotter database

Kaplan-Meier (K-M) plotter (<http://kmplot.com/analysis/>) includes gene chip and RNA-seq data from the GEO, EGA, and TCGA databases, which has been reported to assess the prognostic effect of 54k genes on the patient outcomes among the 21 cancer types, including lung cancers (n = 3,452). To explore the prognostic role of E2F family genes in LUAD and LUSC, the associations between E2F mRNA expression and the overall survival (OS) and recurrence free survival (RFS) was evaluated by the Kaplan-Meier plotter.

Protein expression of E2F family genes in LUAD and LUSC

The immunohistochemistry staining was examined according to the manufacturer's instructions. The 5~8 μm slides were prepared from paraffin-embedded tissues fixed in 4% PFA at 4°C overnight. The slides were heated at 75°C for 30 min, and then 5% FBS was used to reduce unspecific staining. The prepared slides were stained with anti-E2F1, E2F2 and E2F7 antibodies at 4°C overnight. The slides were washed in PBS and incubated with the Goat to rabbit secondary antibodies for 2 hours at room temperature. Histopathological estimates were confirmed by two pathologists at The Second Hospital of Tianjin Medical University based on the WHO classification.

The single-cell RNA expression of E2F family in TISCH

For the single cell RNA-seq (scRNA-seq) expression of lung cancer dataset, the pre-analyzed results of the dataset have been shown in the Tumor Immune Single Cell Hub (TISCH) including the UMAP overview, gene expression, GSEA and GO enrichments [25]. The UMAP plots the entire cells colored by the clusters and cell-type annotations. The malignancy level and major-lineage level were chosen based on the cell-type annotations. Besides, the interested gene expression visualization can be indexed and compared simultaneously based on the current TISCH database. The expression level of E2F family genes at the single-cell resolution were reflected in the UMAP plots. To identify the functions of distinct cell-type populations and gene expressions, the gene set enrichment analysis (GSEA) was performed according to the rank of E2F family genes based on the fold-change from the differential analysis.

Ethics Statement

This study was approved by the Ethics Committee of The Second Hospital of Tianjin Medical University according to the principles of the Declaration of Helsinki. All the datasets were retrieved from the published literatures and online databases. The human LUAD and LUSC paraffin sections were obtained from The Second Hospital of Tianjin Medical University. All the written informed consents were obtained.

Results

The mRNA expression levels of E2F family genes among the divers cancer types

To further evaluate the mRNA expression levels of E2F family genes, we compared the differential expression between the tumor and adjacent normal tissues across all the cancer types by utilizing the TIMER database. As shown in Figure 1, the elevated expression of E2F family members was observed in various kinds of cancers including LUAD and LUSC, which indicated the crucial role of E2F family in tumor progression. Especially, in the Skin Cutaneous Melanoma (SKMC), the mRNA expression of E2F3,

E2F4, E2F5 and E2F8 in the metastatic lesions were significantly increased in contrast to the primary tumors. However, the expression levels of some ones were noticeably reduced in some cancers, such as the expression of E2F2 in rectum adenocarcinoma esophageal carcinoma (READ) and the expression of E2F6 in kidney chromophobe (KICH). In addition, the E2F7 expression was found to be significantly higher in almost all kinds of cancerous tissues compared with the adjacent normal tissues. Importantly, all the E2F family members were almost elevated in LUAD and LUSC with significant differences driving us to investigate the functions in the treatment for lung cancer. Similarly, in the lung cancer statistics of The Cancer Genome Atlas, E2F1 was reported to be overexpressed in LUAD and LUSC compared with the normal tissues evidenced by the fold change of 1.067 and 1.097, respectively (Table 1). The higher expression of E2F2 was found in LUAD (fold change, 1.039) versus normal samples. In the Gao's study, E2F2 was also suggested as one elevated expression factor and one important prognostic factor in the NSCLC [26].

E2F3 expression level was noticeably upregulated in LUAD and LUSC with a fold change of 1.075 and 1.283 (Table 1). Based on the TCGA database, higher expressed E2F5/E2F6/E2F7 was found in the lung carcinomas. A similar trend of E2F5 expression is showed in the Stearman's study [27]. Taken together, the expression of E2F family genes was increased in most of cancers.

Genetic alterations and mRNA expression of E2F family in LUAD and LUSC

To determine the genetic variations of E2F family members in human LUAD and LUSC, several studies from TCGA were included by utilizing the cBioPortal online tool. The mutations were more frequently observed in LUAD, whereas the amplifications appeared more in the LUSC (Figure 2A). This analysis revealed the varying degrees of genetic alteration among member genes of E2F family, of which E2F5 presented the highest incidence rate of genetic variations with the rate of 3% alteration (Figure 2A). The abovementioned statements suggested that the incidences of genetic alteration varied across the E2F family members.

To detect the transcriptomic expression of E2F family genes in LUAD or LUSC and normal lung tissues, transcript per million (TPM) and mRNA expression levels were analyzed by utilizing the Gepia database. The TPM levels were almost elevated in the cancer group, except that were decreased in the E2F4 (Figure 2B). In the LUAD, the expression levels of E2F1 and E2F8 were increased with the significant differences, as well as in the LUSC, we found the elevated expression of E2F1, E2F2, E2F7 and E2F8 (Figure 2B and 2C). Furthermore, the expression of E2F family genes in the disorder stages were also detected suggesting that the E2F2, E2F4, E2F5 and E2F8 groups significantly varied (Figure 3). Amongst them, E2F2 and E2F8 were significantly higher in the later stages of the disease, which indicated that these two factors could be considered as the key markers to label the tumor progression. Collectively, based on the above analyses, E2F8 actually plays a critical role in the lung cancer.

Having identified the mRNA expression levels of E2F1, E2F2 and E2F8, the protein levels of E2F expression were examined by immunohistochemistry in LUAD and LUSC tissues and their counterparts (Figure 4). As shown in Figure 4, we observed that E2F1, E2F2 and E2F8 proteins were expressed at the higher degrees in the cancer tissues than in the normal control.

The correlation between E2F family genes and survival in LUAD and LUSC

To explore the critical efficiency of E2F family genes in the outcomes of patients with LUAD and LUSC, survival plot module of Kaplan-Meier Plotter database was excavated to analyze the correlation between the mRNA levels of E2F family and the recurrence-free survival or the overall survival. For the recurrence-free survival, the overexpression of E2F3 and E2F6 was actually associated with the improved outcomes of LUAD patients. The overexpression of E2F2 impaired the survival of patients with LUSC (Figure 5). For the overall survival, the high level of E2F6 improved the outcome of LUAD patients. Notably, the results revealed that the increased levels of E2F1, E2F4 and E2F8 were significantly associated with the poorer prognosis of LUAD patients (Figure 5). All of the E2F family members produced no significances on the prognosis of LUSC patients. The prognostic potential of E2F family genes in non-small cell lung cancer was further verified by the RNA-seq data of TCGA database.

E2F family genes are correlated with immune infiltration levels in LUAD and LUSC.

Tumor microenvironment (TME) consists of tumor cells and the surrounding environmental ingredients, which presents the high heterogenous features. Several studies have revealed that genes expressed by tumor cells affect the immune cells infiltration levels, which can further play the important roles in the tumor progression. Herein, we explored the correlation between the levels of CD4+ and CD8+ T cells infiltration and the expression of E2F family genes in LUAD and LUSC by analyzing the TIMER database. As shown in Figure 6, we found that the mRNA expression levels of E2F family members were positively related to the degree of tumor purity in LUAD and LUSC. Tumor purity, as an important factor, can influence the analysis between mRNA expression of genes and immune infiltration in tumor samples. Most of E2F family members were obviously positive correlated with infiltrating level of CD4+ and CD8+ T cells. However, the overexpression of E2F2 and E2F5 negatively correlated to the infiltration of CD4+ and CD8+ T cells in both LUAD and LUSC. In the LUSC patient group, E2F6 and E2F7 were also found to impaired the infiltration of CD4+ and CD8+ T cells (Figure 6). These findings strongly suggested that E2F family members play a specific role in immune infiltration, especially those of CD4+ and CD8+ T cells.

The abovementioned results suggested that E2F8 might specifically affect the lung cancer progression. Herein, to further investigate the association between E2F8 expression and various immune infiltrating cells, we analyzed the correlations between mRNA expression and the marker sets of various immune

cells, included CD8+ T cells, CD4+ T cells, Th1 and Th2 cells, Tfh cells, Th17 cells, Tregs cells, exhausted T cells, B cells, monocytes, TAMs, M1 and M2 macrophages, neutrophils and dendritic cells in LUAD and LUSC (Table 2). The results of correlation adjusted by tumor purity revealed that the E2F8 expression level was significantly correlated with most immune cells in LUAD and LUSC. Interestingly, the infiltration of anti-tumor T cells such as CD8+ and CD4+ cells were inhibited by E2F8. On the contrary, Treg cells, TAMs, or exhausted T cells were promoted to infiltrate in the TME. These results indicated that E2F8 may play a vital role in immune escape in the lung cancer microenvironment.

To evaluate the effects of immune cell infiltration on the prognosis of non-small cell lung cancer, the OS analysis of immune infiltration levels was performed including CD4+ T cells, CD8+ T cells, NK cells, Treg cells, B cells, macrophages and cancer-associated fibroblasts based on the TIMER database. The results showed that high levels of CD4+ T cells, Treg cells and macrophages would lead to the worse outcomes of patients (Figure 7).

The function of E2F family genes in LUAD and LUSC progression.

As previously reported, EGFR plays the crucial role in the lung cancer progression and Ki-67 works as an important mark to label the proliferation. Thus, we investigated the correlation of E2F family members with EGFR and Ki-67 expression using the Gepia online tool. As shown in Figure 8A and 8B, EGFR positively correlated with E2F2, E2F4, E2F6~8 and Ki-67 expression was significantly associated with all the E2F family members, which suggested that E2F family genes mostly contributed to the proliferation of tumor cells. To detect the connections with other functional molecules, the network was constructed by analyzing the E2Fs family genes and the 50 top frequently altered neighbor genes. We observed that the cell cycle-related specific genes (CCNE1, CCNE2, CCNF, CDK2 and CDC7), proliferation-related genes (PCNA), epigenetic alteration associated genes (EZH2 and FBXO5) were closely associated with E2F family clusters (Figure 8C).

To identify the signaling pathways associated with E2F family genes differentially activated in LUAD and LUSC, GO enrichment and Gene set enrichment analysis (GSEA) were conducted to obtain potential biological process enriched in non-small cell lung cancer with high expression of E2F family genes. Similar to the molecule networks, the GO enrichments indicated that the cellular growth associated pathways were significantly activated in LUAD and LUSC, such as Cell cycle, DNA replication, ATPase activity, Mitotic Nuclear Division and Catalytic Activity on RNA (Figure 8D). All gene sets that positively or negatively associated with E2F family genes were shown in Table 3. The result of GSEA revealed that immune progresses were positively correlated with E2F family gene expression in LUAD and LUSC, including the INFLAMMATORY_RESPONSE, ALLOGRAFT_REJECTION, IL6_JAK_STAT3_SIGNALING, and HALLMARK_INTERFERON_ALPHA_RESPONSE pathway. Therefore, these results further confirmed the findings that E2F family genes were specifically correlated with tumor cell proliferation and immune infiltration in LUAD and LUSC.

The scRNA expression of E2F family genes in non-small cell lung cancer

To analyze the molecular features of malignant and non-malignant cells in the NSCLC, we obtained 12,193 single cells (GSE143423) from NSCLC based on the TISCH database. To classify the major cell types in the tumor, we performed UMAP analysis and identified seven major clusters: malignant cells, CD8 T cells, Endothelial cells, Monocytes/macrophages, Oligodendrocytes, Pericytes and plasma (Figure 9A). To further explore the E2F family gene expression in each cluster, the used random forest6 analysis was used to identify that E2Fs were mainly expressed in the malignant cells, of which E2F2 and E2F4 stated at the highest levels (Figure 9B). By analyzing differentially expressed genes among the clusters, GO enrichment revealed that “E2F targets”, “G2M checkpoints” and “PI3K-AKT-MTOR” signaling pathways were activated in the malignant cells (Figure 9C), which suggested that E2F family genes in malignant cells significantly correlated to the cell proliferation to promote tumor progression.

Discussion

Lung cancer remains the most common cause of cancer death in males and the second most common cause in females, with an estimated 1.8 million deaths (18%) [28]. About 17.4% of patients can survive more than five years following the initial diagnosis in the United States [7]. However, the average outcomes are much worse in the developing regions with the percentage of about 10.6%. A part of cases is diagnosed at about seventy years old [29].

The histopathology and molecular pathology have been applied in the diagnosis and classification [30], which is important to guide the management and prognostic evaluation. The prognosis estimates and therapeutic strategies depend on the stages and the subgroup of lung cancer and the mutations of specific genes [31], such as the epidermal growth factor receptor (EGFR) or the anaplastic lymphoma kinase (ALK) [32, 33]. NSCLC is still the leading cause of cancer related mortality around the world. The three main subtypes of NSCLC include LUAD, LUSC and large-cell carcinoma [4]. The 5-year survival rate varies from 4 to 17%, relying on the cancer stage and geographical location [34]. The previous studies always took all the subtypes into consideration, while in the current study we separated them based on the subgroup to measure the gene expression and outcome information of LUAD and LUSC, respectively. In spite of the advanced improvement in diagnostic and treatment management, the indicative biological indexes for diagnosis at early stage and correlation with the clinicopathological features should be further detected.

In the previous studies, the functions of E2F transcription factors have been reported in lots of cancers [13]. Even though the role in the tumorigenesis and prognosis has been partially identified, further bioinformatic studies in the overall perspective should be performed. Especially, in the current study, we not only focused on the gene expression and prognostic values of E2F family members, but also investigated the association with immune cell infiltration and the scRNA-seq profiling to highlight the role of E2F family genes in affecting the tumor microenvironment. It has been reported that E2F family genes

play the critical functions in the regulating the cell cycle, such as the G1/S transition in mammalian [35]. Based on the microarray analysis, the transcriptional targets of E2F members included the cyclin proteins, CDKs, checkpoints mediators, DNA repair and replication regulators [36, 37]. However, the unique database of targeted promoters indicated that each member has a unique role during the cell cycle period [38].

Generally, according to the functions, E2F family genes are divided into two groups: transcript activators and suppressors [11]. Amongst them, E2F1, E2F2 and E2F3a were considered as the activators, whereas the six others act as the suppressing factors, including E2F4~8. The activators mainly promote the cell cycle, while the suppressors inhibit the cell cycle and proliferation. But this is not absolute, it's also affected by other factors. For example, the upregulation of E2F8 promoted tumor cell growth in breast cancer by modulating the G1/S phase transition [39]. In the current study, we also found that E2F8 was significantly associated with the poorer prognosis of LUAD patients and E2F6/7 impaired the infiltration of CD4+ and CD8+ T cells. E2F family genes share the similar domains. The DP1,2 heterodimerization domains were detected in the E2F1-6 which allowed the transcription factors to bind to DP1/2. Consecutively, binding to the DP1/2 provided a second DNA binding site to enhance the E2F binding stability [40]. E2F8 binding to the pRB gene has been shown to mask the transactivation domain for transcription activation [39]. For the suppressor E2F4 and E2F5, binding to p107 and p130 mediated the recruitment of repression complexes to silence the targeted gene expression [41].

Another important aspect of the current study is that E2F family genes play an important role in recruitment and regulation of immune infiltrating cells [42]. Our results demonstrate that there is a positive relationship between E2F8 expression level and infiltration level of macrophages and DCs, and significantly positive correlation between infiltration level of CD8+ T, CD4+ T cells and E2F1/2/7 expression in LUAD and LUSC. In addition, the signaling pathways associated with E2F family genes in NSCLC by GSEA revealed that immune progress was positively correlated with expression. Cancer immunotherapy via immune checkpoint blockade depends on releasing T cells from energy and functional exhaustion to eliminate tumor cells [43]. The process of T cell recovery demands the cooperation of the tumor microenvironment, the battlefield between the immune system and tumor, where T cells can be activated persistently to attack the tumor cells [44]. In this battlefield, the interactions between the tumor cells and infiltrating immune cells are critical for the tumor response to immune checkpoint blockade. Notably, the infiltrating immune cells consist of various lymphocytes that can attack the tumor cells or help the tumor cells survive the host immune system [45]. Previously, we reported that PD-L1/PD1 were highly expressed in NSCLC and combination of pembrolizumab and 125I attenuates the aggressiveness of this disorder [46]. Both PD-L1/PD1 and E2F family genes could construct the immunosuppressive microenvironment.

In the current study, the gene expression and prognostic estimation of E2F family members in LUAD and LUSC were comprehensively analyzed by utilizing the online servers, which provided a thorough knowledge of the heterogeneity of lung cancer and the molecular characterization of E2F family members. Our findings revealed that E2F1, E2F2 and E2F8 might promote the tumor growth and

aggressiveness, which was evidenced by the overexpression at the bulk and single-cell resolution and poor prognosis prediction. E2F4 and E2F6 might be considered as the potential outcome markers for the improved treatment of LUAD and LUSC. The significant positive correlation with immune cell infiltration suggested that E2F1, E2F3, E2F7 and E2F8 potentially affected the microenvironmental establishment, which could act as the therapeutic targets in the future.

Declarations

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

Wang S and Guan Z conceived and designed the experiments. Wang S performed the experiments. Zhang J and Meng F performed the pathological analysis of tissues. Wang S and Yan Y contributed to the analysis and interpretation of data. Wang S and Zhang B wrote the manuscript. Guan Z revised the manuscript for intellectual content.

References

1. Bade, B.C. and C.S. Dela Cruz, *Lung Cancer 2020: Epidemiology, Etiology, and Prevention*. Clin Chest Med, 2020. **41**(1): p. 1–24.
2. Nasim, F., B.F. Sabath, and G.A. Eapen, *Lung Cancer*. Med Clin North Am, 2019. **103**(3): p. 463–473.
3. Jones, G.S. and D.R. Baldwin, *Recent advances in the management of lung cancer*. Clin Med (Lond), 2018. **18**(Suppl 2): p. s41-s46.
4. Gridelli, C., et al., *Non-small-cell lung cancer*. Nat Rev Dis Primers, 2015. **1**: p. 15009.
5. Warren, G.W. and K.M. Cummings, *Tobacco and lung cancer: risks, trends, and outcomes in patients with cancer*. Am Soc Clin Oncol Educ Book, 2013: p. 359–64.
6. Bossé, Y. and C.I. Amos, *A Decade of GWAS Results in Lung Cancer*. Cancer Epidemiol Biomarkers Prev, 2018. **27**(4): p. 363–379.
7. Hirsch, F.R., et al., *Lung cancer: current therapies and new targeted treatments*. Lancet, 2017. **389**(10066): p. 299–311.
8. Mairinger, T., *[Histology, cytology and molecular diagnostics of lung cancer]*. Pathologe, 2019. **40**(6): p. 649–661.
9. Woodard, G.A., K.D. Jones, and D.M. Jablons, *Lung Cancer Staging and Prognosis*. Cancer Treat Res, 2016. **170**: p. 47–75.
10. Reck, M., et al., *Updated Analysis of KEYNOTE-024: Pembrolizumab Versus Platinum-Based Chemotherapy for Advanced Non-Small-Cell Lung Cancer With PD-L1 Tumor Proportion Score of 50%*

- or Greater. *J Clin Oncol*, 2019. **37**(7): p. 537–546.
11. Sladky, V.C., et al., *E2F-Family Members Engage the PIDDosome to Limit Hepatocyte Ploidy in Liver Development and Regeneration*. *Dev Cell*, 2020. **52**(3): p. 335-349.e7.
 12. Osswald, M., et al., *Brain tumour cells interconnect to a functional and resistant network*. *Nature*, 2015. **528**(7580): p. 93–8.
 13. Kent, L.N. and G. Leone, *The broken cycle: E2F dysfunction in cancer*. *Nat Rev Cancer*, 2019. **19**(6): p. 326–338.
 14. Whitfield, T.W., et al., *Functional analysis of transcription factor binding sites in human promoters*. *Genome Biol*, 2012. **13**(9): p. R50.
 15. Sozzani, R., et al., *Interplay between Arabidopsis activating factors E2Fb and E2Fa in cell cycle progression and development*. *Plant Physiol*, 2006. **140**(4): p. 1355–66.
 16. Sun, C.C., et al., *Comprehensive Analysis of the Expression and Prognosis for E2Fs in Human Breast Cancer*. *Mol Ther*, 2019. **27**(6): p. 1153–1165.
 17. Trikha, P., et al., *E2f3 in tumor macrophages promotes lung metastasis*. *Oncogene*, 2016. **35**(28): p. 3636–46.
 18. Hurst, C.D., et al., *Inactivation of the Rb pathway and overexpression of both isoforms of E2F3 are obligate events in bladder tumours with 6p22 amplification*. *Oncogene*, 2008. **27**(19): p. 2716-27.
 19. Lu, Z., et al., *E2F-HDAC complexes negatively regulate the tumor suppressor gene ARHI in breast cancer*. *Oncogene*, 2006. **25**(2): p. 230–9.
 20. Bellmunt, J., *Stem-Like Signature Predicting Disease Progression in Early Stage Bladder Cancer. The Role of E2F3 and SOX4*. *Biomedicines*, 2018. **6**(3).
 21. Gruszka, R., et al., *mRNA and miRNA Expression Analyses of the MYC/E2F/miR-17-92 Network in the Most Common Pediatric Brain Tumors*. *Int J Mol Sci*, 2021. **22**(2).
 22. Yang, S., et al., *Distinct expression pattern and prognostic values of pituitary tumor transforming gene family genes in non-small cell lung cancer*. *Oncol Lett*, 2019. **18**(5): p. 4481–4494.
 23. Tang, Z., et al., *GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses*. *Nucleic Acids Res*, 2017. **45**(W1): p. W98-w102.
 24. Li, T., et al., *TIMER2.0 for analysis of tumor-infiltrating immune cells*. *Nucleic Acids Res*, 2020. **48**(W1): p. W509-w514.
 25. Sun, D., et al., *TISCH: a comprehensive web resource enabling interactive single-cell transcriptome visualization of tumor microenvironment*. *Nucleic Acids Res*, 2021. **49**(D1): p. D1420-d1430.
 26. Gao, Z., et al., *Expression and prognostic value of E2F activators in NSCLC and subtypes: a research based on bioinformatics analysis*. *Tumour Biol*, 2016. **37**(11): p. 14979–14987.
 27. Stearman, R.S., et al., *Analysis of orthologous gene expression between human pulmonary adenocarcinoma and a carcinogen-induced murine model*. *Am J Pathol*, 2005. **167**(6): p. 1763–75.
 28. Sung, H., et al., *Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries*. *CA Cancer J Clin*, 2021.

29. Yatabe, Y., et al., *Best Practices Recommendations for Diagnostic Immunohistochemistry in Lung Cancer*. J Thorac Oncol, 2019. **14**(3): p. 377–407.
30. Rodriguez-Canales, J., E. Parra-Cuentas, and Wistuba, II, *Diagnosis and Molecular Classification of Lung Cancer*. Cancer Treat Res, 2016. **170**: p. 25–46.
31. Shi, J., et al., *Somatic Genomics and Clinical Features of Lung Adenocarcinoma: A Retrospective Study*. PLoS Med, 2016. **13**(12): p. e1002162.
32. da Cunha Santos, G., F.A. Shepherd, and M.S. Tsao, *EGFR mutations and lung cancer*. Annu Rev Pathol, 2011. **6**: p. 49–69.
33. Chapman, A.M., et al., *Lung cancer mutation profile of EGFR, ALK, and KRAS: Meta-analysis and comparison of never and ever smokers*. Lung Cancer, 2016. **102**: p. 122–134.
34. Aberle, D.R., et al., *Reduced lung-cancer mortality with low-dose computed tomographic screening*. N Engl J Med, 2011. **365**(5): p. 395–409.
35. Thwaites, M.J., M.J. Cecchini, and F.A. Dick, *Analyzing RB and E2F during the G1-S transition*. Methods Mol Biol, 2014. **1170**: p. 449–61.
36. Roskoski, R., Jr., *Cyclin-dependent protein serine/threonine kinase inhibitors as anticancer drugs*. Pharmacol Res, 2019. **139**: p. 471–488.
37. Markey, M.P., et al., *Unbiased analysis of RB-mediated transcriptional repression identifies novel targets and distinctions from E2F action*. Cancer Res, 2002. **62**(22): p. 6587–97.
38. Pierce, A.M., et al., *Differential activities of E2F family members: unique functions in regulating transcription*. Mol Carcinog, 1998. **22**(3): p. 190–8.
39. Ye, L., et al., *Upregulation of E2F8 promotes cell proliferation and tumorigenicity in breast cancer by modulating G1/S phase transition*. Oncotarget, 2016. **7**(17): p. 23757–71.
40. Lukas, J., et al., *Deregulated expression of E2F family members induces S-phase entry and overcomes p16INK4A-mediated growth suppression*. Mol Cell Biol, 1996. **16**(3): p. 1047–57.
41. Chen, H.Z., S.Y. Tsai, and G. Leone, *Emerging roles of E2Fs in cancer: an exit from cell cycle control*. Nat Rev Cancer, 2009. **9**(11): p. 785–97.
42. Liao, P., S. Han, and H. Qu, *Expression, Prognosis, and Immune Infiltrates Analyses of E2Fs in Human Brain and CNS Cancer*. Biomed Res Int, 2020. **2020**: p. 6281635.
43. He, X. and C. Xu, *Immune checkpoint signaling and cancer immunotherapy*. Cell Res, 2020. **30**(8): p. 660–669.
44. Rezvani, K. and R.H. Rouse, *The Application of Natural Killer Cell Immunotherapy for the Treatment of Cancer*. Front Immunol, 2015. **6**: p. 578.
45. Bremnes, R.M., et al., *The Role of Tumor-Infiltrating Lymphocytes in Development, Progression, and Prognosis of Non-Small Cell Lung Cancer*. J Thorac Oncol, 2016. **11**(6): p. 789–800.
46. Wang, S., et al., *Combination of pembrolizumab and (125)I attenuates the aggressiveness of non-small cell lung cancer*. Oncol Lett, 2020. **19**(6): p. 4142–4150.

The mRNA expression of E2F family genes among diverse human cancer types based on the TIMER database. The boxplots showing the expression levels of E2F family genes. The red bars for the tumor samples and blue bars for the adjacent normal samples. *p < 0.05, **p < 0.01, ***p < 0.001.

Adrenocortical carcinoma (ACC), bladder urothelial carcinoma (BLCA), breast invasive carcinoma (BRCA), cervical squamous cell carcinoma and endocervical adenocarcinoma (CESC), cholangiocarcinoma (CHOL), colon adenocarcinoma (COAD), rectum adenocarcinoma esophageal carcinoma (READ), the lymphoid neoplasm diffuse large B-cell lymphoma (DLBC), esophageal carcinoma (ESCA), glioblastoma multiforme (GBM), glioma (GBMLGG), head and neck squamous cell carcinoma (HNSC), kidney chromophobe (KICH), kidney renal clear cell carcinoma (KIRC), kidney renal papillary cell carcinoma (KIRP), acute myeloid leukemia (LAML), brain lower-grade glioma (LGG), liver hepatocellular carcinoma (LIHC), lung adenocarcinoma (LUAD), lung squamous cell carcinoma (LUSC), mesothelioma (MESO), ovarian serous cystadenocarcinoma (OV), pancreatic adenocarcinoma (PAAD), pheochromocytoma and paraganglioma (PCPG), prostate adenocarcinoma (PRAD), sarcoma (SARC), skin cutaneous melanoma (SKCM), stomach adenocarcinoma (STAD), stomach and esophageal carcinoma (STES), testicular germ cell tumors (TGCT), thyroid carcinoma (THCA), thymoma (THYM), uterine corpus endometrial carcinoma (UCEC), uterine carcinosarcoma (UCS) and uveal melanoma (UVM).

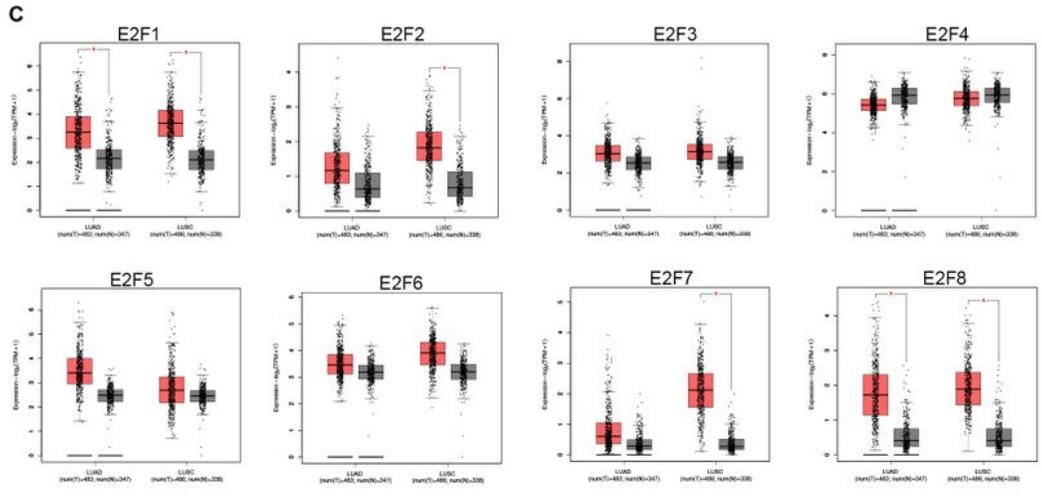
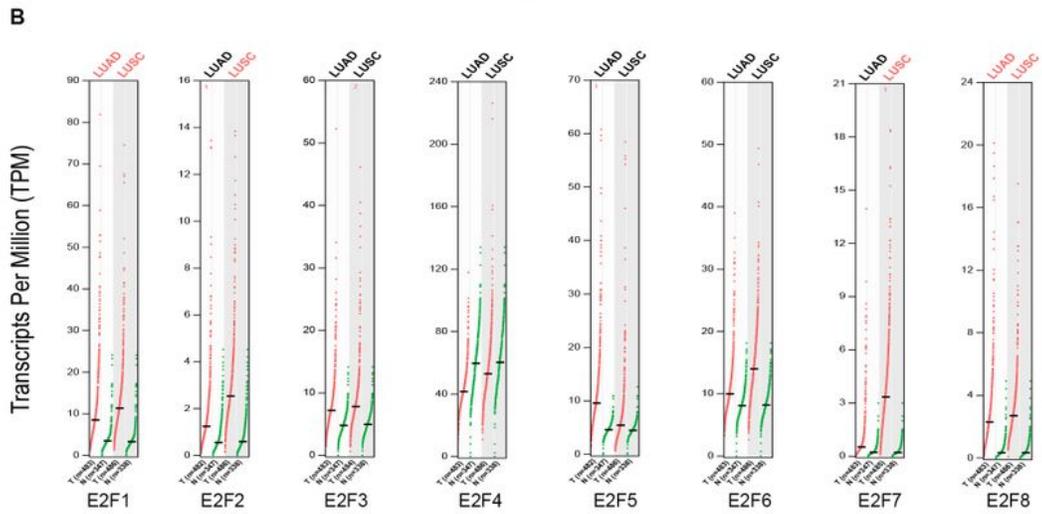
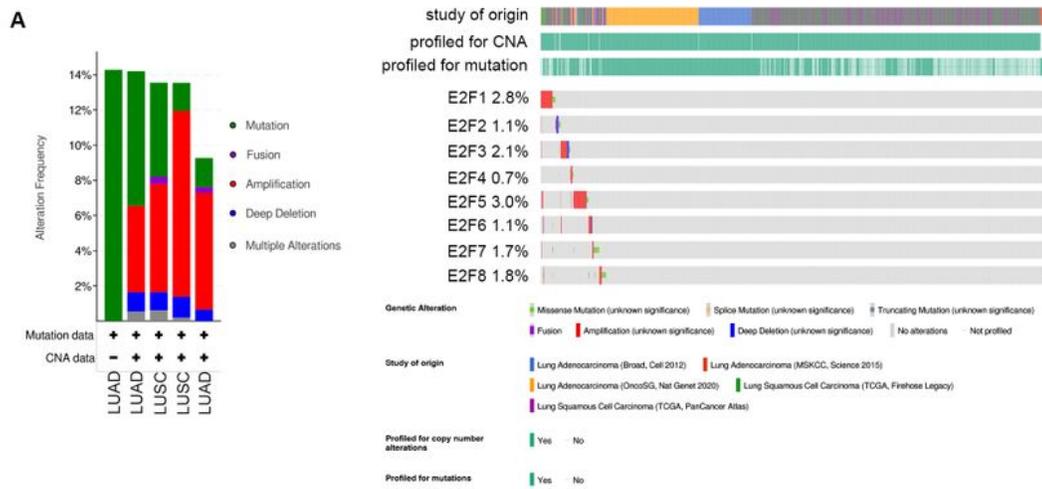


Figure 2

Genetic variations and mRNA levels of E2F family genes in LUAD and LUSC. A. The genomic alteration of E2F family including copy number missense mutation with unknown significance (green), gene fusion (purple), amplification (red) and deep deletion (blue) were analyzed in LUAD and LUSC based on the cBioportal database. B. The transcripts per million of E2F family genes in LUAD and LUSC based on the

Gepia database. C. The expression of E2F family genes in the cancer and normal tissues of LUAD and LUSC based on the Gepia database.

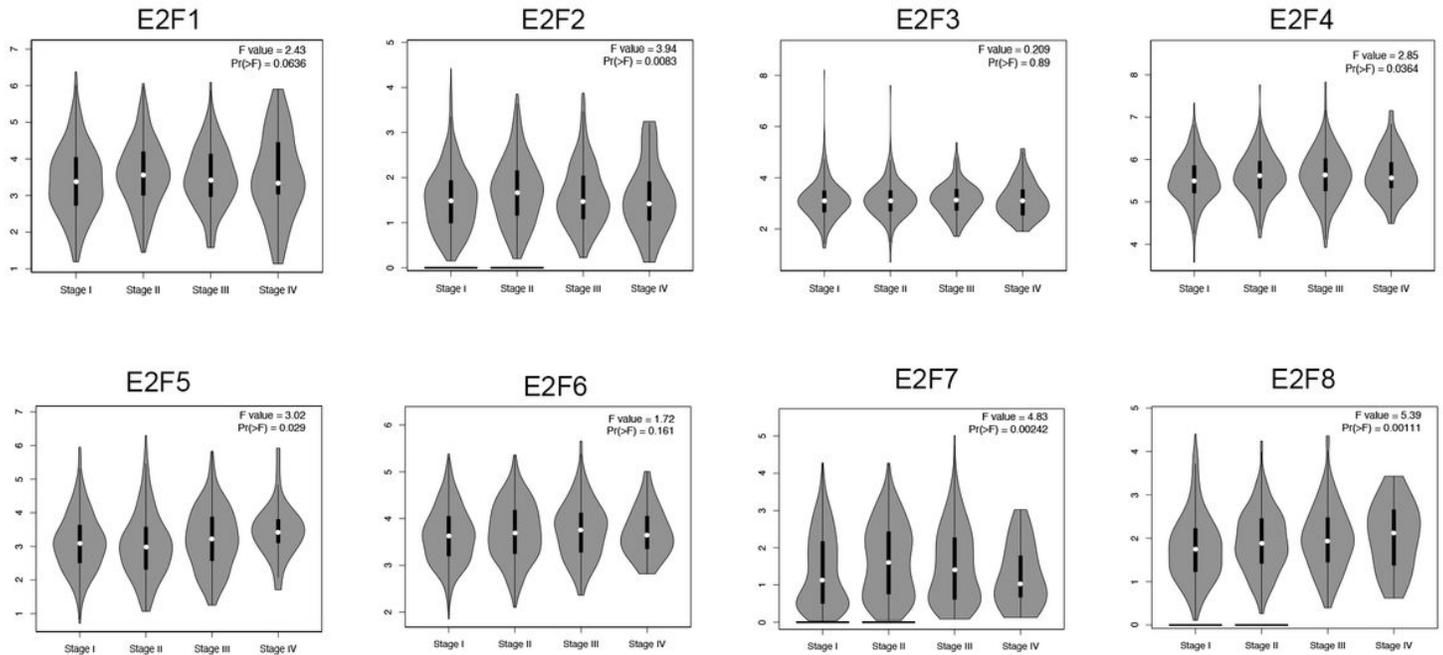


Figure 3

The correlation between E2F family gene expression and tumor stage in LUAD and LUSC based on the Gepia database.

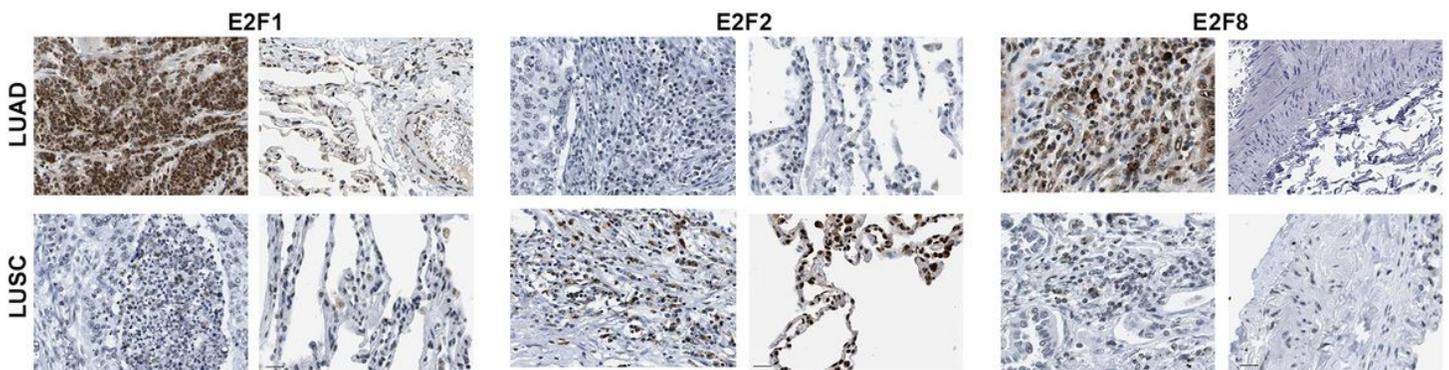


Figure 4

The expression of E2F1, E2F2 and E2F7 in LUAD and LUSC examined by immunohistochemistry.

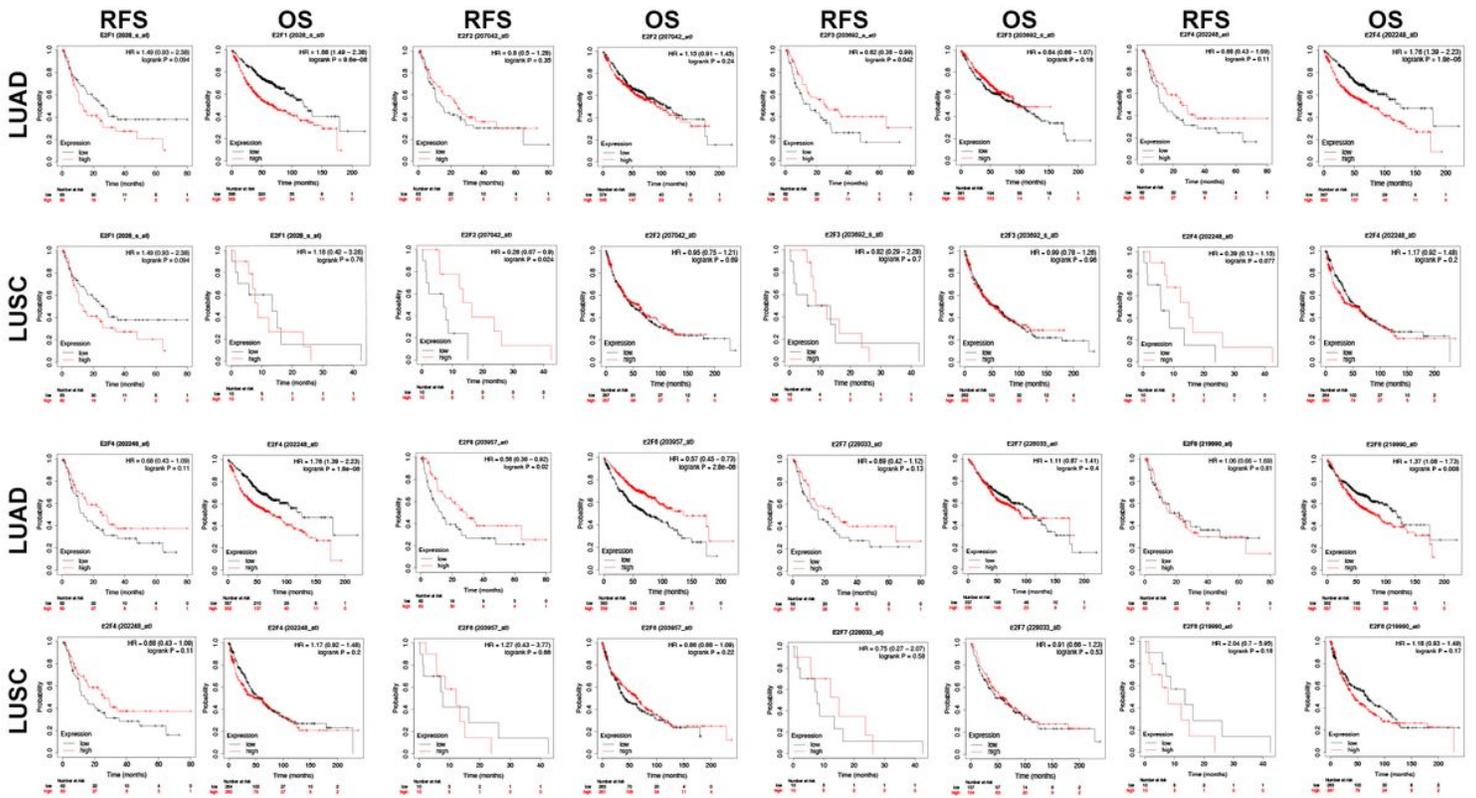


Figure 5

The association between E2F family genes and prognosis of LUAD and LUSC patients. The OS and RFS analysis of E2F family genes was performed using Kaplan–Meier Plotter online tool.

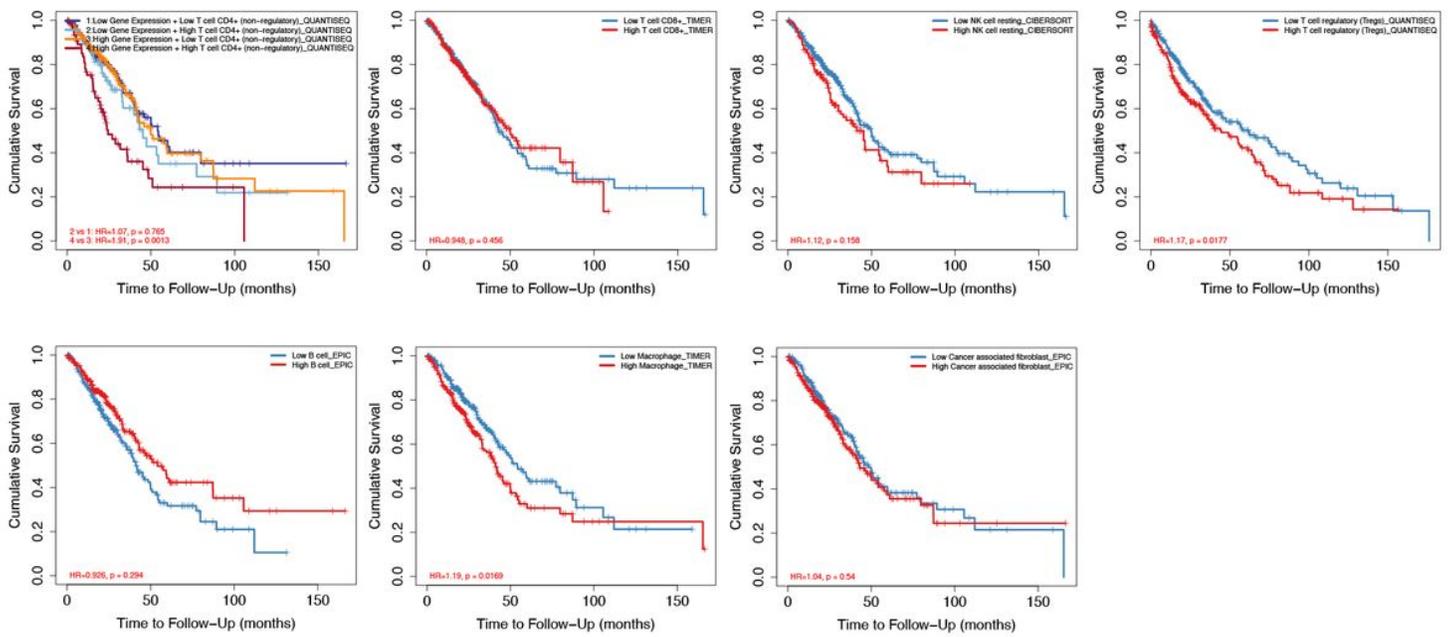


Figure 7

The association between prognosis of non-small cell lung cancer patients and immune infiltration levels. The OS analysis of immune infiltration levels was performed including CD4+ T cells, CD8+ T cells, NK cells, Treg cells, B cells, macrophages and cancer-associated fibroblasts based on the TIMER database.

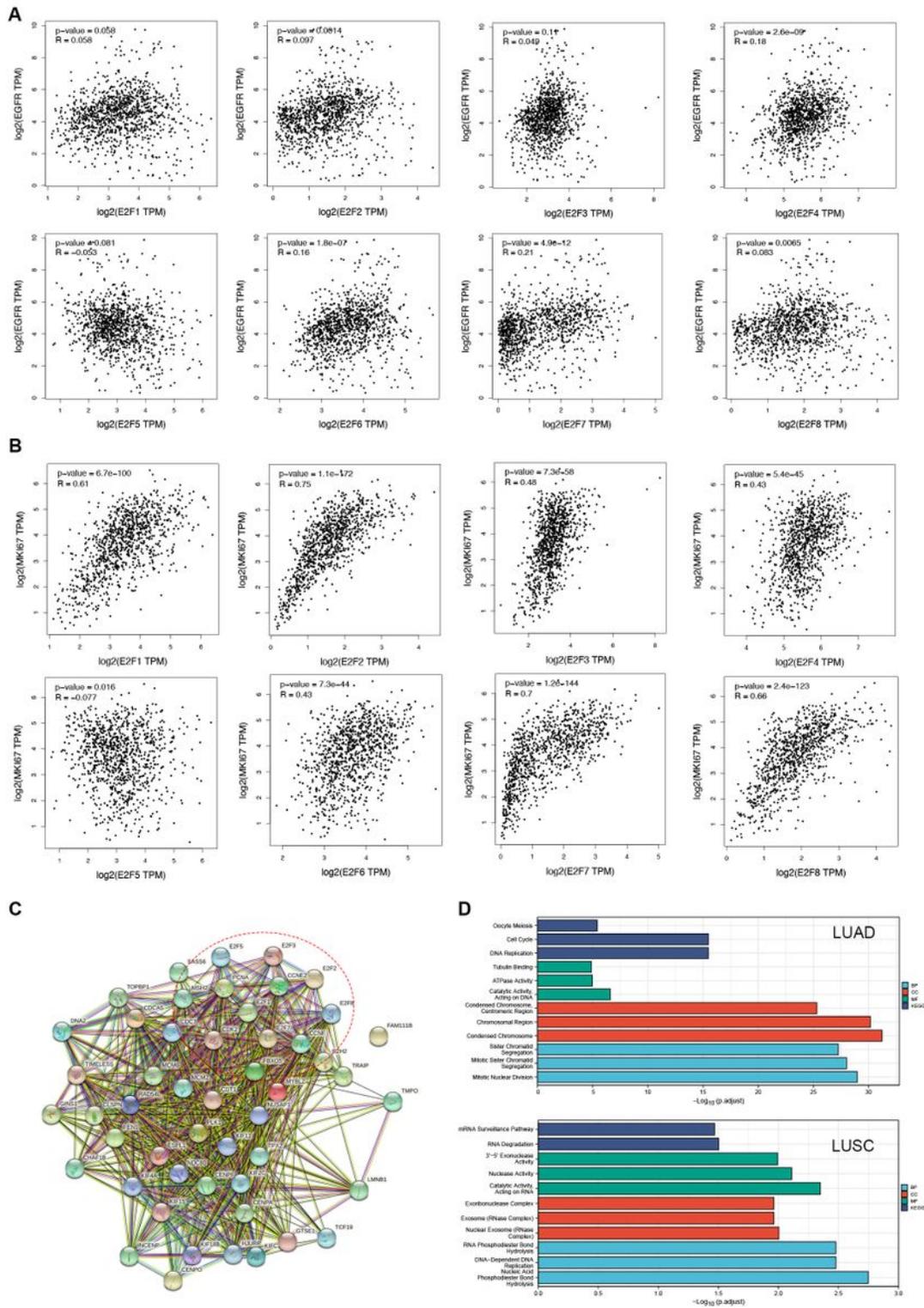


Figure 8

The association between the expression of E2F family genes and other functional factors in LUAD and LUSC. A. The Pearson analysis was performed between E2F family genes and EGFR expression based on the Gepia database. B. The Pearson analysis was performed between E2F family genes and MKI67 expression based on the Gepia database. C. The network for E2F family genes and the 50 most frequently altered neighbor genes. D. GO enrichment analysis for E2F family genes in LUAD and LUSC.

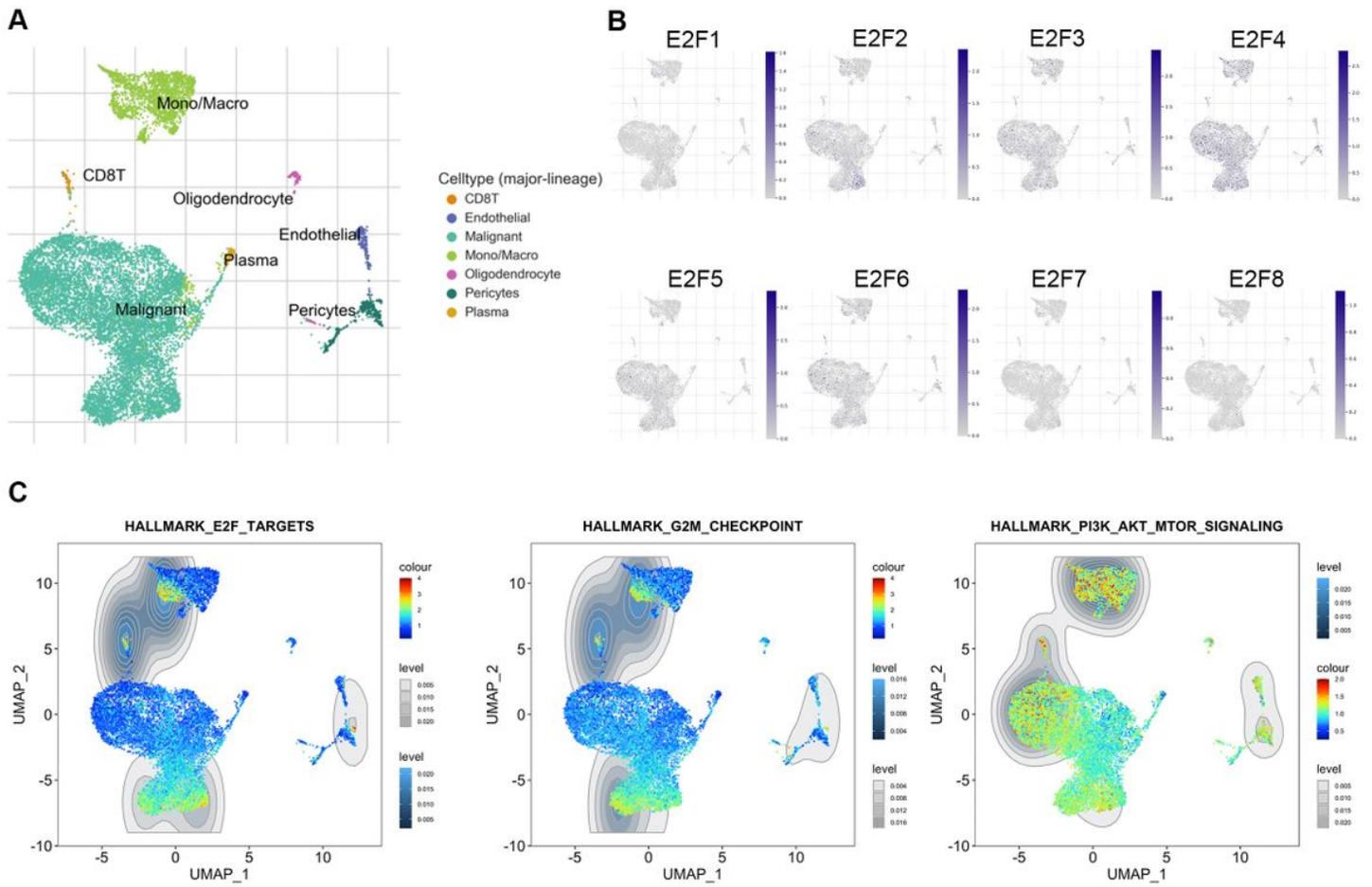


Figure 9

Single-cell RNA sequence analysis of non-small cell lung cancer. A. UMAP analysis of 23,892 cells of non-small cell lung cancer. B. The expression of E2F family genes in the UMAP at the single cell resolution. C. The activated signaling pathway based on the expression of E2F family genes in the UMAP.