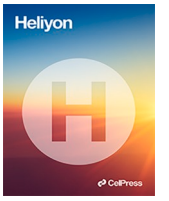
Heliyon 10 (2024) e33928

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon



Research article

Single cell transcriptomic analysis reveals tumor immune infiltration by NK cells gene signature in lung adenocarcinoma

Yimin Zhu, Xiuhua Wu, Yunjiao Zhang, Jie Gu, Rongwei Zhou, Zhong Guo *

Department of Pulmonary and Critical Care Medicine, Shanghai Sixth People's Hospital Affiliated to Shanghai Jiao Tong University School of Medicine, Shanghai, China

ARTICLE INFO

Keywords:
Single-cell RNA-Sequencing
Lung adenocarcinoma (LUAD)
NK gene signature (NKGS)
Prognosis
TGF-β signaling
Immunosuppressive

ABSTRACT

Background: Natural Killer (NK) cells are vital components of the innate immune system, crucial for combating infections and tumor growth, making them pivotal in cancer prognosis and immunotherapy. We sought to understand the diverse characteristics of NK cells within lung adenocarcinoma (LUAD) by conducting single-cell RNA sequencing analyses.

Methods: Using the scRNA-seq dataset for multiple primary lung cancers (MPLCs), we examined two major NK cell groups, NK1 and NK2, comparing the expression profiles of 422 differentially expressed NK signature genes. We identified eight genes (SPON2, PLEKHG3, CAMK2N1, RAB27B, CTBP2, EFHD2, GOLM1, and PLOD1) that distinguish NK1 from NK2 cells. A prognostic signature, the NK gene signature (NKGS) score, was established through LASSO Cox regression. High NKGS scores were linked to poorer overall survival in TCGA-LUAD patients and consistently validated in other datasets (GSE31210 and GSE14814).

Results: Functional analysis revealed an enrichment of genes related to the TGF-β signaling pathway in the high NKGS score group. Moreover, a high NKGS score correlated with an immunosuppressive tumor microenvironment (TME) driven by immune evasion mechanisms. We also observed reduced T-cell receptor (TCR) repertoire diversity in the high-risk NKGS group, indicating a negative association between inflammation and risk score.

Conclusion: This study introduced the innovative NKGS score, differentiating NK1 from NK2 cells. High NKGS scores were associated with the TGF-β pathway and provided insights into LUAD prognosis and immune activities.

1. Introduction

Non-small cell lung cancer (NSCLC) is a predominant subtype of lung cancer, posing a formidable global health challenge due to its widespread prevalence and high mortality rates [1]. Remarkably, NSCLC constitutes an overwhelming 85 % of all lung cancer cases, underscoring its clinical significance and amplifying the call for advanced therapeutic solutions [2,3]. One of the inherent complexities of NSCLC is its marked genetic and phenotypic heterogeneity. Such diversity not only complicates diagnosis for clinicians but also impedes the creation of universally effective treatment approaches [4,5]. As a result, the spotlight in modern NSCLC research has shifted towards personalized precision therapy [6,7]. In the complex landscape of this disease, biomarkers have become essential tools, holding potential for predicting patient outcomes and refining treatment approaches [8–10].

* Corresponding author.
E-mail address: zhongguo0819@163.com (Z. Guo).

https://doi.org/10.1016/j.heliyon.2024.e33928
Received 3 November 2023; Received in revised form 28 June 2024; Accepted 30 June 2024
Available online 2 July 2024

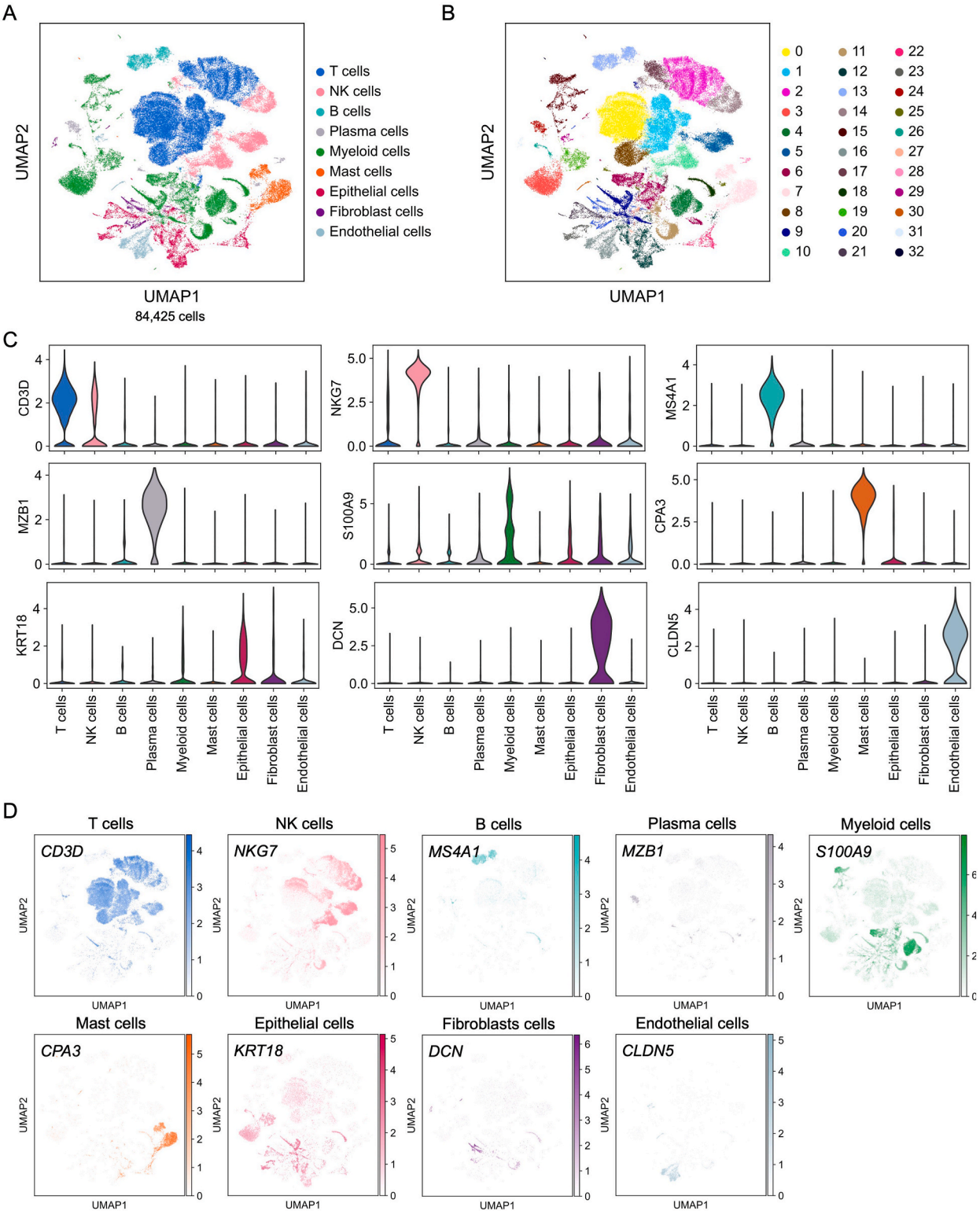


Fig. 1. Comprehensive analysis of MPLCs single cell atlas. (A) UMAP plots of nine major cell lineages (T cells, NK cells, B cells, Plasma cells, Myeloid cells, Mast cells, Epithelial cells, Fibroblast cells, and Endothelial cells) from MPLCs patients. (B) UMAP plots of 84,425 cells, colored by 33 cell clusters. (C) Violin plot for expression of one canonical marker gene in each major cell lineage. (D) UMAP plots of nine major cell lineages, colored by corresponding canonical marker gene.

In recent years, biomarkers such as PD-L1 expression and tumor mutation burden (TMB) have been frequently utilized in the clinical arena, primarily to anticipate the response to immunotherapy, a promising line of treatment that has offered considerable benefits for various malignancies [11–13]. However, these markers do not encapsulate the entire story. The tumor microenvironment (TME) of NSCLC is varied and dynamic, suggesting that relying solely on current biomarkers might not provide a comprehensive understanding of the potential therapeutic outcomes [14–16]. The significance of biomarker research in NSCLC can hardly be overstated. Biomarkers offer insights into the biological behavior of tumors, paving the way for more accurate diagnostic and prognostic tools [16]. Moreover, by identifying specific molecular and cellular features, biomarkers can guide the selection of targeted treatments, enhancing therapeutic efficacy and potentially reducing adverse side effects [17–19]. However, they often successfully predict the therapeutic outcome of immunotherapy only for a subset of LUAD (Lung Adenocarcinoma) patients, indicating a pressing need for expanded biomarker discovery [20–23]. Therefore, a contemporary research hotspot in NSCLC encompasses the development of novel predictive models. This approach is expected to revolutionize therapeutic strategies, by not only improving prognosis accuracy but also by refining treatment modalities, thus aiming to enhance patient outcomes and survival rates [20–23].

Natural Killer (NK) cells, have a profound connection to malignancies, especially lung cancer, a predominant driver of cancer-induced fatalities worldwide [24–26]. Leading this frontier is the adoptive NK cell therapy are administered back to patients [27–29]. These fortified NK cells demonstrate augmented endurance and robust anti-tumor capabilities within the challenging, immunosuppressive landscape of lung cancer's tumor microenvironment (TME) [30,31]. Compounding this challenge, factors such as TGF-β and IL-10, secreted by the tumor cells, serve to dampen NK cell efficacy significantly [32–35]. Add to this the incursion of immunosuppressive cohorts like MDSCs and Tregs, and NK cell activity is further stifled. Disturbingly, preliminary findings indicate that lung tumors may actively induce NK cell apoptosis, depleting their numbers within the TME [36–39]. This potent immunosuppression presents formidable obstacles in treatment efficacy. While frontline modalities like chemotherapy or radiation might successfully target and eliminate cancer cells, an attenuated immune backdrop may falter in maintaining lasting tumor oversight, ushering in potential relapses [40]. In summary, the critical defense role of NK cells against lung cancer is undeniable; however, their compromised cytotoxic prowess in impacted individuals amplifies the urgency for pioneering therapeutic interventions [27].

The advent of single-cell RNA-sequencing (scRNA-seq) technology, coupled with sophisticated data analysis techniques, has unlocked unparalleled potential to decipher the molecular nuances of various immune cell subsets within the TME [41–43]. Previous studies have reported that scRNA-seq datasets can serve as a compelling approach to forecast patient prognosis and their responsiveness to immunotherapy [44–46]. Specifically for NSCLC, scRNA-seq has been adeptly employed to probe alveolar macrophages, an integral pillar of the lung's immunological architecture [47–50]. In this study, we performed an exhaustive analysis of scRNA-seq data from multiple primary lung cancers (MPLCs) [51]. Our aim was to delve into the molecular intricacies of tumor-infiltrating NK cells and pinpoint distinctive marker genes associated with these NK cells. Moreover, we devised a Natural Killer gene signature (NKGS) score as a prognostic tool for LUAD by integrating scRNA-seq and TCGA-LUAD RNA-seq analyses. We further validated the predictive accuracy of the NKGS using two independent cohorts from the Gene Expression Omnibus (GEO) database. Additionally, we delved into the interplay between the NKGS and the immunosuppressive tumor microenvironment (TME) and conducted a potential functional analysis within the LUAD context.

2. Results

2.1. Cellular atlas of multiple primary lung cancers (MPLCs)

To understand the detailed cellular heterogeneity within the tumor microenvironment (TME) of lung cancer, a reanalysis was conducted on the scRNA-seq dataset GSE200972 [51]. This dataset encompasses 19 lung samples (Figure S1A), with 5 tumor samples from the inferior lobe (TI), 4 tumor samples from the middle lobe (TM), and 2 tumor samples from the superior lobe (TS). Additionally, there are 4 normal tissue samples adjacent to TI (NI), 3 normal tissue samples adjacent to TM (NM), and 1 normal tissue sample adjacent to TS (NS). For data processing, a principal component analysis (PCA) with n_comps = 50 was performed, followed by the construction of a neighborhood graph emphasizing the most variable genes and setting n_neighbors at 10. Resultant cell clusters were annotated using CellTypist [52] and canonical cell markers, revealing 84,425 cells categorized into nine primary cell subsets (Fig. 1A). Using the Louvain algorithm (resolution = 0.4) via Scanpy [53], 33 unique cell clusters were discerned (Fig. 1B). Expectedly, these major cell lineages demonstrated consistency across samples based on canonical marker gene expression (Fig. 1C–S1B, and S1C). Cells like T cells, NK cells, B cells, Plasma cells, myeloid cells, mast cells, epithelial cells, fibroblast cells, and endothelial cells displayed prominent characterization through high expressions of *CD3D*, *NKG7*, *MS4A1*, *MZB1*, *S100A9*, *CAP3*, *KRT18*, *DCN*, and *CLDN5*, respectively (Fig. 1C and D). Interestingly, each cell cluster showed enriched expression of its distinguishing canonical marker genes, indicating shared expression patterns within clusters and clear distinctions between cell lineages (Figure S1D). In conclusion, we described the landscapes of cell clusters across 19 MPLCs samples for further analysis.

2.2. Characterization of different NK cell clusters

NK cells are crucial components of the innate immune system that play vital roles in tumor immunosurveillance. NK cells are typically categorized based on their surface expression of CD56 and CD16. The two primary subsets are: NK1 ($CD56^{dim}CD16^{hi}$) and NK2 ($CD56^{bright}CD16^{lo}$). In lung cancer, the tumor microenvironment can impact NK cell functionality. Tumors often develop mechanisms to suppress NK cell activity, such as producing immunosuppressive cytokines, upregulating inhibitory ligands, or recruiting regulatory cells that inhibit NK cell functions.

Following stringent quality control measures and filtering, a total of 9099 NK cells were identified and further categorized into two distinct subsets (Fig. 2A). These NK cells were organized into seven unique clusters derived from NI, TI, NM, and TN samples (Fig. 2B). It was noted that NK1 cells exhibited enhanced expression of CD16 (*FCGR3A*), *PRE1*, and *GZMH*, indicated their identity as the $CD56^{dim}CD16^{hi}$ NK cell phenotype (Fig. 2C and D, S2A, and S2B) [54]. In contrast, NK2 cells demonstrated heightened expression levels of *VIM*, *FOS*, and *JUN* (Fig. 2C and D, S2A, and S2B). The data further emphasized that most cellular markers exhibited subset-specific expression patterns (Figure S2C and S2D). Recognized for their potent cytotoxic capabilities, NK1 cells are instrumental in the direct elimination of tumor cells. A reduction in the NK1 cell proportion within tumor samples implies a potential compromise in the anti-tumor defenses in these zones. It's well-observed that tumors can foster an immunosuppressive milieu, thereby limiting the

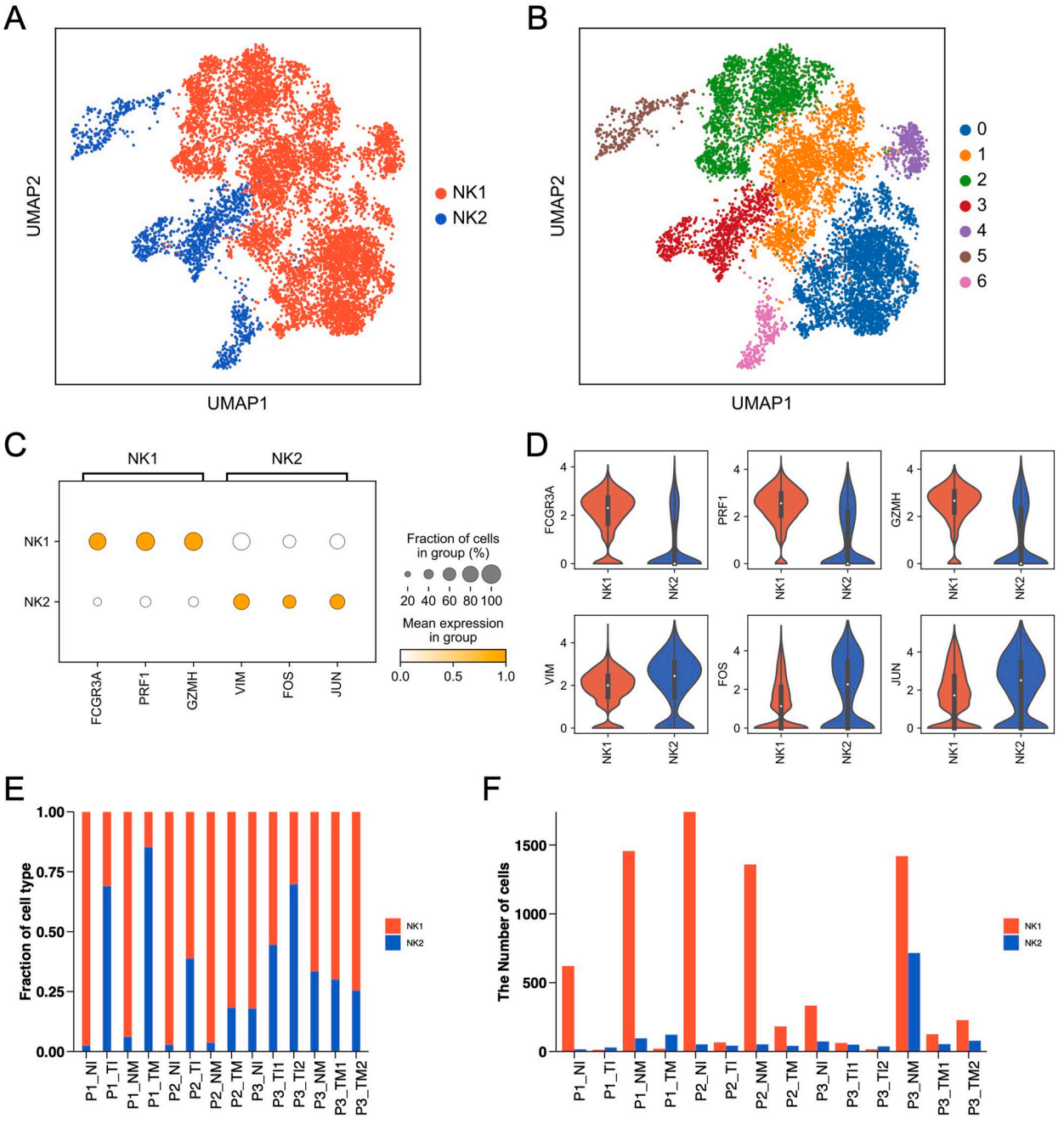
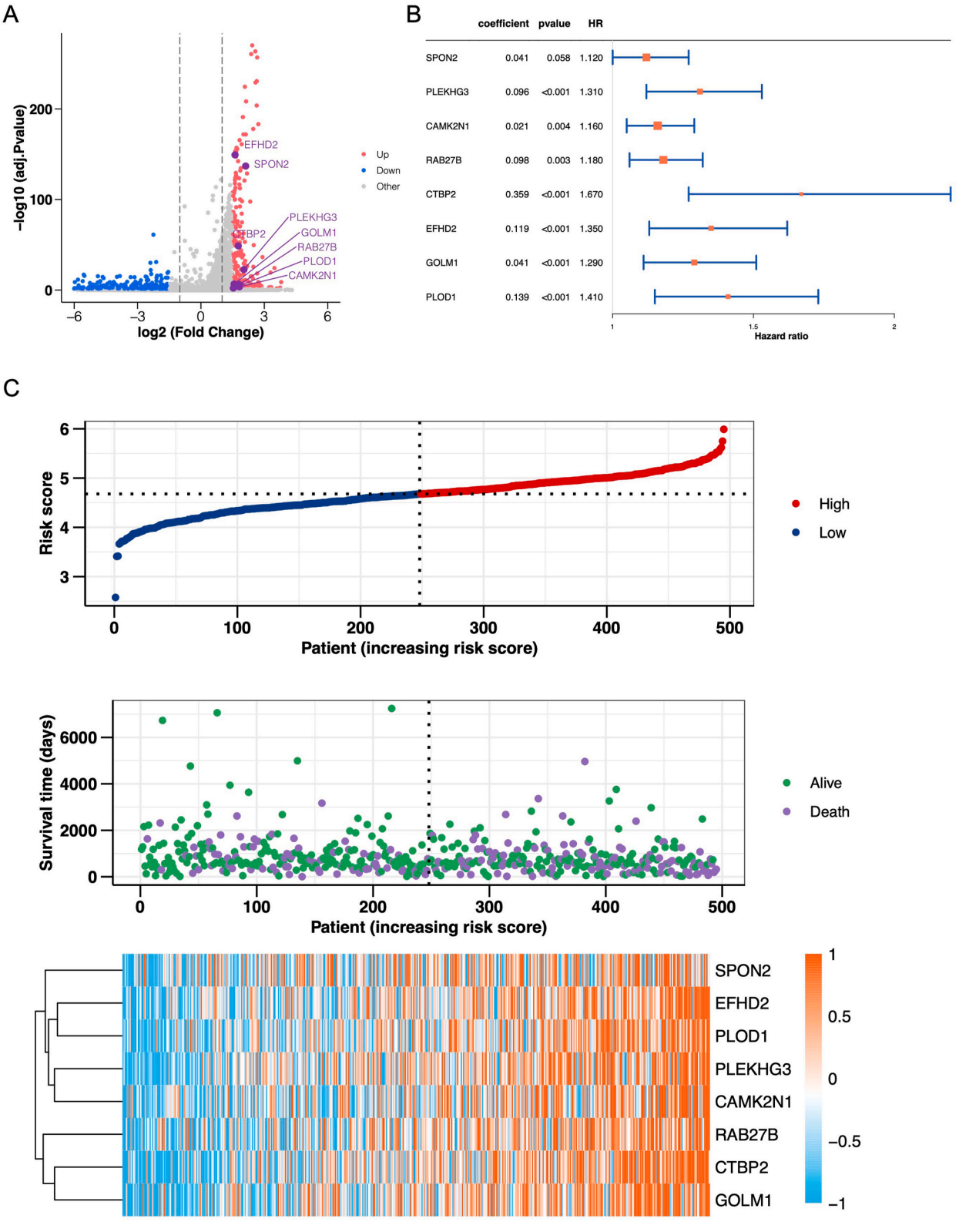


Fig. 2. Expression genes of NK cell clusters in MPLCs patients. (A) UMAP plot of 9099 NK cells (NK1, $CD56^{dim}CD16^{high}$; NK2, $CD56^{bright}CD16^{low}$), colored by two cell types. (B) UMAP plot of NK cells, colored by seven cell clusters. (C) Dot plot of three represented expressed genes in NK1 and NK2 cell types. (D) Violin plot for expression of three canonical marker genes (NK1, *FCGR3A*, *PRF1*, and *GZMH*; NK2, *VIM*, *FOS*, and *JUN*) in each NK cell type. (E) The proportion of each sample in two NK cell types. (F) The number of cells in two NK cell types.



(caption on next page)

Fig. 3. Construction of prognostic model with NK signature genes in the TCGA-LUAD cohort. (A) Volcano plot for expression of differentially expressed genes between NK1 and NK2 cell types. (B) Kaplan–Meier curves of survival analysis compared the overall survival of TCGA-LUAD patients between NKGS-high and NKGS-low groups. (C) ROC curves of the NKGS for predicting the risk of death at 1, 3, and 5 years. (D) The distribution of risk score (top), survival status (middle), and expression (bottom) of the identified eight NK cell marker genes. (E) Validation of the NKGS in GSE31210 (n = 226) cohort. (F) Validation of the NKGS in GSE14814 (n = 114) cohort.

recruitment, activation, and longevity of cytotoxic immune responders, NK1 cells included. A prominent presence of NK1 cells in adjacent normal tissues suggests that the immunosuppressive effects of the tumor might be largely restricted to its immediate surroundings. Collectively, these insights propose that the ratio of NK1 to NK2 cells could offer significant diagnostic or prognostic value.

2.3. Construction of the prognostic signature based on NK cell signature genes

To assess the functionality within NK cell clusters, differentially expressed genes (DEGs) for each cluster were determined by contrasting NK1 and NK2 cell types using the Wilcoxon–Mann–Whitney test. This analysis yielded 218 up-regulated and 203 down-regulated genes (Fig. 3A). Aiming to construct a prognostic signature from these DEGs, the TCGA-LUAD cohort was employed as a training set. Subsequently, a least absolute shrinkage and selection operator (LASSO) Cox proportional hazards regression analysis was executed, culminating in the selection of eight most predictive genes for the risk prognostic model. Risk score = (0.0413 × *SPON2* expression) + (0.119 × *EFHD2* expression) + (0.139 × *PLOD1* expression) + (0.096 × *PLEKHG3* expression) + (0.021 × *CAMK2N1* expression) + (0.098 × *RAB27B* expression) + (0.359 × *CTBPO* expression) + (0.041 × *GOLM1* expression) (Fig. 3B). Furthermore, an overview displaying the distribution of risk scores, survival statuses, and gene expression profiles was provided (Fig. 3C). This overview underscored a higher mortality rate in the high-risk group, which also presented elevated gene expression levels. Therefore, we defined a novel prognostic NK gene signature (NKGS) for further analysis.

Setting the risk score threshold at 4.67805, achieved by ranking risk scores in descending order, patients were segmented into low-risk (n = 247) and high-risk (n = 248) cohorts. Kaplan–Meier (KM) analysis illustrated that high-risk patients exhibited markedly diminished overall survival (OS) compared to their low-risk counterparts (Fig. 4A, HR = 1.76, $P = 0.00016$). To gauge the predictive accuracy of this risk prognostic model, time-dependent ROC curves for OS were analyzed, revealing 1-, 3-, and 5-year AUC values of 0.624, 0.659, and 0.641, respectively (Fig. 4B).

2.4. Validation of the NKGS in different independent cohorts

In this study, two independent GEO cohorts (GSE31210 and GSE14814) were incorporated to ascertain the reliability of the NKGS score. Initially, using a consistent formula, the risk scores of each patient across the four GEO cohorts were computed. Subsequently, by ranking the risk scores in descending order, the median value was employed to segregate patients into either the high-risk or low-risk category. Kaplan–Meier (KM) analysis was then conducted on both independent GEO cohorts. Results consistently revealed that the high-risk group exhibited a poorer prognosis compared to the low-risk group across both datasets: GSE31210 (Fig. 4C, HR: 2.38, 95 % CI: 1.36–4.15, $P = 0.014$) and GSE14814 (Fig. 4D, HR: 2.18, 95 % CI: 1.44–3.29, $P = 0.0016$). Collectively, these outcomes underscore the potential of the NKGS risk score as a predictive tool for the prognosis of LUAD patients. To further validate the application of the NKGS score in external cohorts, we analyzed the 5-year and 10-year survival data from TCGA-LUAD, GSE31210, and GSE14814 using KM curve analysis. The outcomes consistently revealed similar trends for both 5-year and 10-year survival durations, underscoring the efficacy of the NKGS score as a prognostic signature for LUAD (Figure S3A-S3F).

Subsequently, an assessment was undertaken to determine if a high-risk score serves as an independent factor associated with OS. A multivariate Cox proportional hazards analysis, incorporating variables such as age, gender, tumor stage, and risk score, demonstrated that the risk score indeed functioned as an independent predictor of OS in the TCGA-LUAD cohort ($P < 0.001$, Fig. 4E). In essence, these results suggest that the risk score can prognosticate the outcomes for LUAD patients independently of variables like age and gender, though it remains interdependent with tumor stage.

2.5. Functional enrichment analysis of the NKGS related genes

To decipher the biological functions and mechanisms associated with the NKGS risk score, we turned to the Gene Set Enrichment Analysis (GSEA). Our goal was to identify pathways predominantly influenced by genes associated with the NKGS score. Utilizing the hallmark gene sets (c2.cp.v7.4.symbols.gmt) as a reference, we observed a pronounced enrichment of the TGF-β signaling pathway in the high-risk group (Fig. 5A). A more detailed GSEA revealed that the differentially expressed genes (DEGs) related to a heightened NKGS score were predominantly aligned with the TGF-β signaling pathway (Fig. 5B, NES = 2.915, $P = 8.77e\text{-}11$). Concurrently, using the Gene Set Variation Analysis (GSVA), we evaluated the functional activity differences across the NKGS risk score groups. This analysis revealed an increased GSVA pathway activity in the NKGS-high group, notably linked to Pan-F-TBRS (Fig. 5C, $P = 1.1e\text{-}05$) [55]. More interestingly, we found that the NKGS-high group was more enriched in the EMT pathway. While the median value for EMT1 was higher, it was not statistically significant (Fig. 5D, $P = 0.27$). However, the signature score values for EMT2 ($P = 0.00024$) and EMT3 ($P = 0.0035$) were significantly elevated (Fig. 5E and F), indicating a potential for activated EMT pathway. Additionally, when analyzing the expression levels of the signature genes for Pan-F-TBRS and EMT, we found higher expression in the NKGS-high group. Most of the signature gene expressions were more enriched in the NKGS-high group, suggesting that the NKGS score might

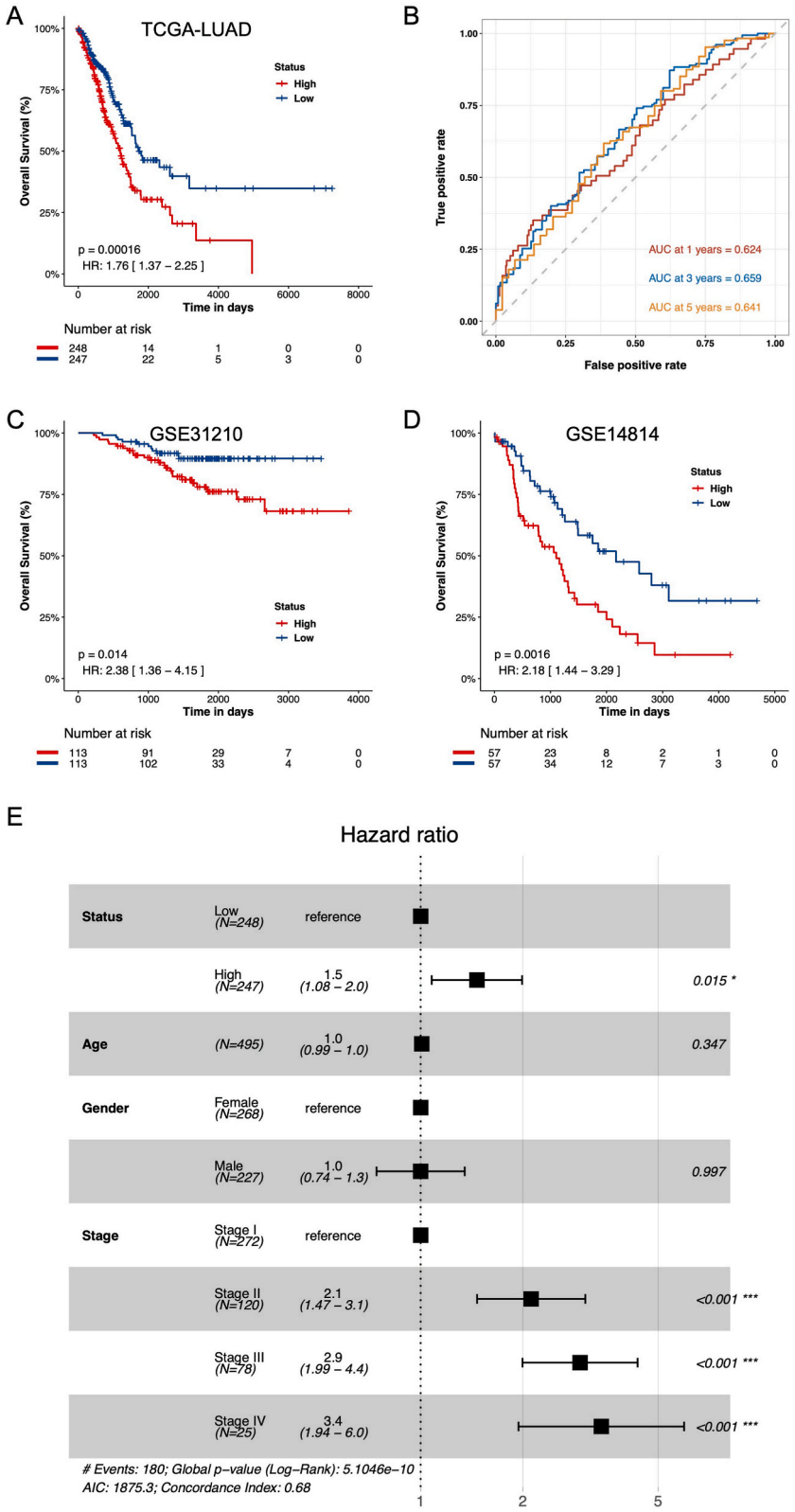


Fig. 4. Cox regression analyses between NKGS-high and NKGS-low groups. (A) Hazard ratios of eight NK signature genes in univariate Cox models that were significantly associated with overall survival. (B) Forest plots show HRs and 95 % confidence intervals (horizontal ranges) derived from cox regression survival analyses for overall survival of eight NK signature genes, including Status, Age, Gender, and Stage.

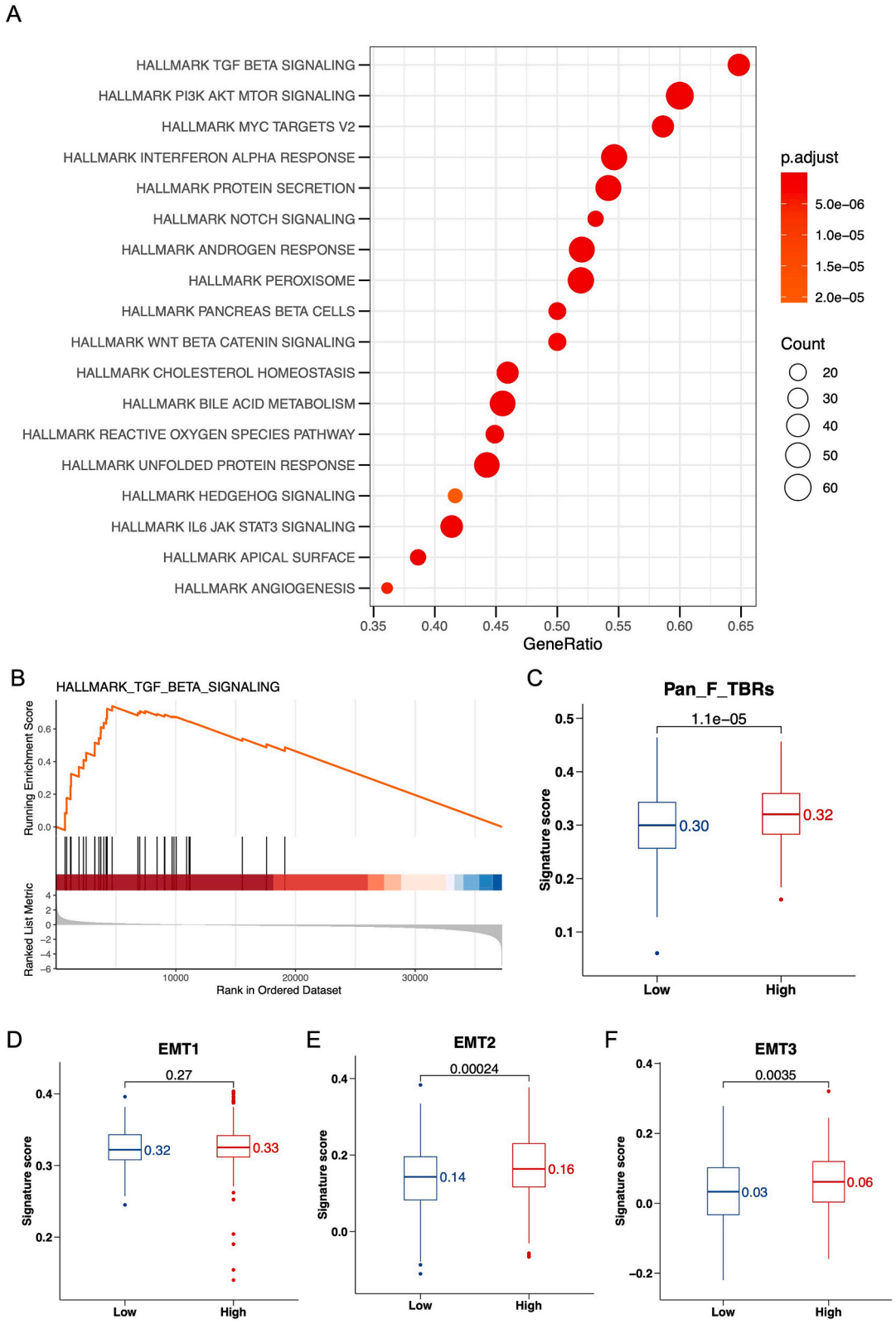


Fig. 5. Function analysis between NKGS-high and NKGS-low groups. (A) Function enrichment of NKGS-high and NKGS-low groups in TCGA-LUAD cohort. (B) GSEA analysis of hallmark TGF-β signaling pathway. GSVA activity analysis of Pan-F-TBRs (C), EMT1 (D), EMT2 (E), and EMT3 (F) pathway.

activate the potential mechanisms of the TGF-β and downstream EMT pathways (Figure S4A-S4D). These observations illuminate the likely attributes of the NKGS-high group, which appears to be defined by enhanced immunosuppression and reduced immune infiltration. Building on this foundation, we aimed to further delineate the disparities in immune infiltration between the NKGS-high and NKGS-low groups.

2.6. Analysis of immune infiltration influenced by the NKGS score

To gain a deeper understanding of the attributes and roles of the NKGS-high and NKGS-low groups, we investigated the relationship between NKGS and immune cell infiltration in TCGA-LUAD patients. Leveraging the Tumor Immune Dysfunction and Exclusion (TIDE) algorithm [56], we gauged the potential for immune escape in the TCGA-LUAD cohort. Our analysis revealed that patients in the NKGS-high category had notably increased Cancer Associated Fibroblasts (CAF) scores (Fig. 6A, $P = 2.3e\text{-}05$), heightened exclusion scores (Fig. 6B, $P = 0.00068$), elevated Myeloid-Derived Suppressor Cells (MDSCs) scores (Fig. 6C, $P = 0.00062$), and a higher TIDE score (Fig. 6D, $P = 0.00068$). These metrics derive from gene signatures that signify immune evasion through T-cell exclusion. In sum, the data suggests that the NKGS-high group exhibits elevated values in the aforementioned indicators, implying these patients may be situated in a more immunosuppressive tumor environment. Such an environment could be linked to poorer clinical trajectories and reduced efficacy of immunotherapeutic interventions.

3. Prediction of immunotherapy benefits in LUAD patients

T cells are instrumental in mounting the tumor immune response, making it possible for immunotherapies to accurately identify and target tumor cells. Expanding on previous studies, T cells can be genetically modified to carry T cell receptors (TCRs) tailored to specific cancer antigens, streamlining the detection of antigens presented through the MHC. TCR analysis has emerged as a meaningful biomarker in evaluating anti-tumor immune activities [57]. In our study, we assessed TCR diversity by referencing the TCR repertoire database [45,58] and drew comparisons between the NKGS-high and NKGS-low groups. Intriguingly, we found that the NKGS-high group had considerably diminished TCR diversity (Fig. 7A, $P = 0.02062$) and a reduced Shannon diversity index (Fig. 7B, $P = 0.02562$) in contrast to the NKGS-low group. This data indicates that individuals with a lower NKGS risk score could be better positioned to benefit from immunotherapy treatments.

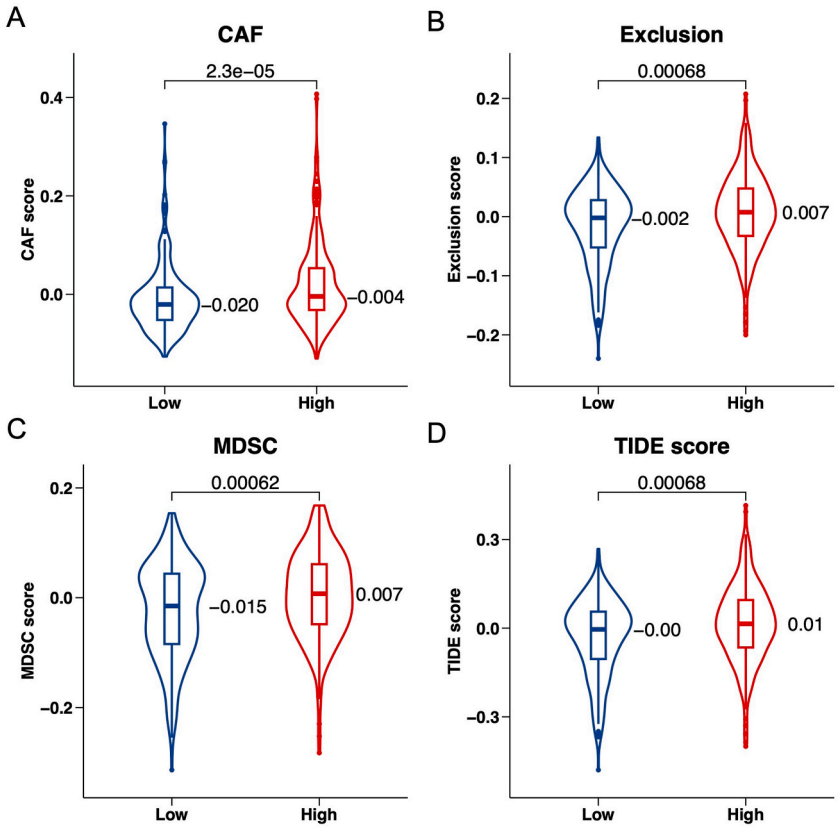


Fig. 6. Tumor microenvironment analysis between NKGS-high and NKGS-low groups. TIDE analysis in NKGS-high and NKGS-low groups, including CAF score (A), Dysfunction score (B), Exclusion score (C), M2 score (D), MDSC score (E), and TIDE score (F).

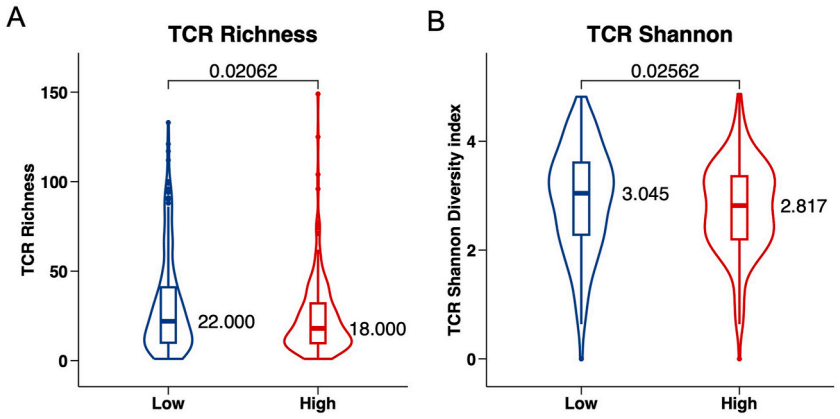


Fig. 7. The role of NKGS in predicting immunotherapeutic benefit. TCR richness (A) and TCR diversity (B) between NKGS-high and NKGS-low groups.

4. Discussion

Exploring the tumor microenvironment of LUAD has opened the door for us to efficiently decode the molecular characteristics and mechanisms in clinical diagnosis and prognosis research. Single-cell RNA sequencing has further deepened our understanding of the cellular heterogeneity in NSCLC. Some studies even suggest that enhancing NK cell activity might bolster the immune response against cancer. In recent years, particularly the immunotherapy strategies based on NK cells, have emerged and shown new hope in lung cancer treatments. These treatment strategies aim to better utilize the anti-tumor functions of NK cells, offering patients more effective treatment options. Single-cell analysis has also revealed the resistance mechanisms to immunotherapy. The TGF-β signaling pathway, closely related to this, plays a crucial role in the initiation of EMT. Thus, targeting this pathway might further enhance the efficacy of immunotherapy. In conclusion, single-cell analysis techniques provide us with a unique insight into the molecular and cellular heterogeneity of NSCLC and help pinpoint potential therapeutic targets for lung cancer. A deeper understanding of NSCLC heterogeneity and exploration of new treatment strategies may offer a brighter therapeutic outlook for patients afflicted with this severe condition.

In this study, we developed a distinct prognostic prediction model anchored on an NK gene signature tailored for LUAD patients in the TCGA database. We further validated this model rigorously using two separate cohorts from the GEO dataset. The NKGS features nine NK cell marker genes, namely *SPON2*, *PLEKHG3*, *CAMK2N1*, *RAB27B*, *CTBP2*, *EFHD2*, *GOLM1*, and *PLOD1*. Most of these genes hold significant associations with LUAD patient prognosis or with NK cell functionality.

Spondin 2 (*SPON2*) is a protein-coding gene that has been studied in the context of various diseases, including cancer. Identifying potential biomarkers for this condition is crucial. Research reveals that *SPON2*, while known to be involved in metastasis and cancer progression, plays a vital role in bone metastasis in lung adenocarcinoma (ADC). Elevated *SPON2* levels enhance the migration, invasion, and epithelial-to-mesenchymal transition of ADC cells, while its suppression hinders these processes. Additionally, *SPON2* levels are higher in metastatic bone tissues than primary ADC tissues, correlating with increased *MMP2* and *MMP9* expressions. Moreover, in T1a stage NSCLC patients, *SPON2* expression is upregulated, facilitated by exosomal *HOTAIRM1* interactions in CAFs, promoting NSCLC cell movement [59,60]. Pleckstrin homology and RhoGEF domain-containing G3 (*PLEKHG3*) is a gene that encodes a protein with a Rho guanine nucleotide exchange factor (RhoGEF) domain, which is involved in activating the Rho family of small GTPases. Research on *PLEKHG3* in cancer, especially in lung cancer, is rarely reported. In this study, we found that *PLEKHG3* can be used as one of the genes in the prognostic model. This is the first time it has been reported in lung cancer prognosis research. Calcium/Calmodulin Dependent Protein Kinase II Inhibitor 1 (*CAMK2N1*), in lung cancer, its high expression has been linked to unfavorable outcomes [61]. *RAB27B* belongs to the RAB family of small GTPases, which are pivotal in vesicle trafficking processes. The heightened *RAB27B* expression may serve as an adverse prognostic indicator for patients with lung SQCC [62]. C-terminal binding protein 2 (*CTBP2*) is a well-established transcriptional corepressor with diverse roles in cellular functions such as development, differentiation, and apoptosis. Utilizing the chi-square test and Kaplan-Meier analysis, it was evident that elevated *CTBP2* expression was associated with a more aggressive tumor phenotype and adverse prognosis. Furthermore, our findings indicate that *CTBP2* knockdown augments NSCLC cell sensitivity to *cis*-diamminedichloroplatinum (CDDP) by inhibiting the Wnt/β-catenin pathway. Collectively, these insights underscore the pivotal role of *CTBP2* in NSCLC progression and CDDP responsiveness, suggesting that targeting *CTBP2* could offer a novel therapeutic avenue for NSCLC management [63]. EF-hand domain-containing protein D2 (*EFHD2*), a protein linked to numerous cellular functions, has been scrutinized in diverse health contexts [64]. Additionally, heightened *EFHD2* expression correlates with increased metastasis and epithelial-mesenchymal transition (EMT), making it a promising independent predictor for postsurgical recurrence in stage I lung adenocarcinoma patients [65]. Golgi membrane protein 1 (*GOLM1*) is a transmembrane glycoprotein found in the Golgi cisternae and has been linked to the carcinogenesis of various cancers. Furthermore, DNA copy number variations and methylation could be key mechanisms driving *GOLM1* dysregulation in LUAD [66]. Evidence is mounting that members of the procollagen-lysine, 2-oxoglutarate 5-dioxygenase (*PLOD*) gene family, namely *PLOD1*, *PLOD2*, and *PLOD3*, play roles in cancer progression and metastasis. Elevated *PLOD* expression significantly correlates with adverse survival outcomes and immune cell

infiltration in LUAD. This suggests that *PLOD* family genes might serve as emerging biomarkers for adverse prognosis and potential targets for LUAD immunotherapy [67].

The differential distribution of NK1 cells ($CD56^{dim}CD16^{hi}$) in tissues suggests a significant immunological distinction between tumor and adjacent normal regions. Specifically, higher concentrations of NK1 cells in normal tissues (NI and NM) adjacent to TI and TM hint at a potential suppressive mechanism within the tumor environment, restricting these cytotoxic cells. This may indicate the tumor's ability to modulate its surrounding microenvironment, evading immune surveillance. Conversely, the consistent or diminished presence of NK2 cells ($CD56^{bright}CD16^{lo}$) across both tumor and surrounding tissues suggests that their primary non-cytotoxic, regulatory functions are not significantly altered by the tumor's presence. Collectively, these patterns underscore the tumor's potential immune-evasive strategies and the intricate interplay between different NK cell subsets in cancer progression. Understanding these distributions could provide insights into therapeutic interventions aimed at harnessing NK cell activity against tumors.

The NKGS prognostic model stands out as a powerful predictor of patient outcomes across both training and validation cohorts. To further understand the underpinnings of the NKGS classifications, we carried out GO pathway and GSEA assessments. Interestingly, the group with a high NKGS score displayed significant involvement in the TGF-β signaling pathway. Through the TIDE algorithms, we discerned diminished immune infiltration and immunosuppression in this group. Additionally, our analysis of the TCR repertoire's complexity and diversity revealed a counter relationship between inflammatory actions and the risk score in the high-risk NKGS cohort.

Furthermore, the elevated exclusion score hints at potential mechanisms that might obstruct effective immune cell infiltration into tumors. Such barriers can be either physical, like a dense extracellular matrix, or biochemical, driven by immunosuppressive cytokines or metabolic alterations. Concurrently, a higher MDSC score reflects an augmented immunosuppressive condition, wherein these cells can suppress T-cell activation, hindering effective anti-tumor immune responses. Lastly, the TIDE score's elevation suggests that tumors in the NKGS-high group might be adept at evading immune responses, possibly rendering them less responsive to conventional immunotherapies. Collectively, these observations intimate that patients in the NKGS-high category may be confronted with more aggressive tumor behavior and could face challenges in leveraging immunotherapies due to a profoundly immunosuppressive microenvironment.

In this study, our mechanistic analyses are largely observational, prompting a need for more in-depth exploration to unravel the link between NKGS gene expression and LUAD prognosis. We have delineated and validated an eight-gene signature rooted in NK cell genes, laying the groundwork for personalized therapeutic strategies for LUAD patients. However, it's vital to approach our results with a clear understanding of their inherent limitations. Future studies should focus on these aspects and further clarify the clinical relevance of the discerned gene signature.

5. Methods and materials

5.1. Collection of single-cell RNA sequencing and GEO datasets

The original scRNA-seq dataset from multiple primary lung cancers (MPLCs) was sourced from GSE200972 [51]. This dataset comprised data from four patients, totaling 19 samples. These included 5 tumor samples from the inferior lobe (TI), 4 tumor samples from the middle lobe (TM), and 2 tumor samples from the superior lobe (TS). Additionally, there were 4 normal tissue samples adjacent to TI (NI), 3 normal tissue samples adjacent to TM (NM), and 1 normal tissue sample adjacent to TS (NS).

Gene expression profiles for lung adenocarcinoma (LUAD) cohorts were procured from the TCGA portal (https://portal.gdc.cancer.gov/). The raw read counts and associated clinical details such as age, gender, stage, overall survival (OS), and vital status for LUAD patients were accessed from the UCSC Xena website (https://xenabrowser.net/datapages/). To further validate the OS status of the proposed gene set, we referenced the microarray data and clinical details from the NCBI GEO database for GSE31210 (n = 226) and GSE14814 (n = 114).

5.2. Single cell analysis and cell clustering

The single-cell dataset matrix was imported into Scanpy (version 1.9.1) [53] for in-depth analysis. Principal component analysis was conducted using 'sc.tl.pca' with parameters set to n_comps = 50. The neighborhood graph was clustered using 'sc.pp.neighbors' with parameters set to n_neighbors = 10. Dimensionality was reduced using highly variable genes, and cell clusters were identified using the Louvain algorithm via the "sc.tl.louvain" module, with the parameter resolution set to 0.4. These cell clusters were subsequently visualized using the uniform manifold approximation and projection (UMAP) module "sc.pl.umap". The annotations for each cell cluster were determined based on the expression of known marker genes using CellTypist (https://github.com/Teichlab/celltypist) [52].

6. Construction and validation of the prognostic signature

A univariate Cox regression analysis was conducted to determine the prognostic value of the marker genes between NK1 and NK2 cells for OS in TCGA-LUAD patients. Genes with a $P < 0.05$ were designated as prognostic genes. These genes were then evaluated using the least absolute shrinkage and selection operator (LASSO) Cox proportional hazards regression via the "glmnet" R package. Following this analysis, a risk model was formulated by linearly combining the mRNA expression of the genes with their corresponding risk coefficients. Nine genes were identified as key candidate prognosis-related genes for further investigation. Patients were stratified

into NKGS-high risk and NKGS-low risk groups based on the top one-third cutoff value. To verify the prognostic significance of the NKGS score, receiver operating characteristic (ROC) curves [68] were constructed, and the area under the curve (AUC) was computed using the "survivalROC" R package to gauge the accuracy of the NKGS model.

7. Survival analysis

Using the "survival" and "survminer" R packages, Kaplan–Meier (KM) curves were employed to analyze the expression and prognosis of NKGS-related genes in TCGA-LUAD. The KM survival curve analysis indicated a correlation between the NKGS-high group and poorer overall survival (OS). The predictive capability of the NKGS was further corroborated through survival analysis on four independent GEO datasets.

7.1. Differential expressed gene analysis

The expression of individual genes in each cluster was compared to the remaining cells using the "sc.tl.rank_genes_groups" module through the Wilcoxon rank sum test. A gene was defined as up-regulated or down-regulated based on a $P < 0.05$, with a log (fold change) ≥ 1 or ≤ -1 serving as the cutoff criteria, respectively.

8. Gene set enrichment analysis (GSEA)

Gene Set Enrichment Analysis (GSEA) was conducted to identify significantly enriched pathways using the hallmark gene sets (c2.cp.v7.4.symbols.gmt) from MSigDB (http://software.broadinstitute.org/gsea/msigdb/).

8.1. Gene set variation analysis (GSVA)

Gene Set Variation Analysis (GSVA) was carried out to discern the activity of enriched pathways between the NKGS-high and NKGS-low groups. The enrichment scores for each gene set in the TCGA-LUAD samples were determined using the ssGSEA algorithm via the "GSVA" R package.

9. TIDE analysis

The Tumor Immune Dysfunction and Exclusion (TIDE) algorithm was utilized to predict potential ICB responses [10]. Patients with elevated TIDE scores are more likely to experience antitumor immune evasion, often correlating with reduced response rates to ICB treatment [60]. Recent studies further underscore its effectiveness in predicting or assessing ICB therapy efficacy [43,69].

9.1. TCR analysis

TCR repertoire analysis is recognized as a valuable biomarker for anti-tumor immune responses. In this research, the TCR richness and diversity across each NKGS group were examined in reference to a previous study [58].

9.2. Statistics analysis

Differences of statistical significance were evaluated using a two-tailed Student's t-test on the R platform (R v4.0.3). Multivariate analysis employing the Cox proportional hazards model was executed using the R packages ("survival", "survminer", and "forestplot") to pinpoint independent factors linked to OS in both TCGA-LUAD and GEO cohorts. The *P* value was corrected using the false discovery rate (FDR), with values or FDR <0.05 deemed significant. An adjusted $P < 0.05$ served as the threshold criterion.

Funding

Not available.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Data availability statement

The single-cell RNA-seq dataset was sourced from the GEO database using the accession number GSE200972. We obtained the

TCGA-LUAD cohort dataset from The Cancer Genome Atlas (TCGA) portal (https://portal.gdc.cancer.gov/). Additionally, we retrieved datasets GSE31210 and GSE14814 from the GEO database.

CRediT authorship contribution statement

Yimin Zhu: Writing – original draft, Visualization, Investigation, Formal analysis, Conceptualization. **Xiuhua Wu:** Resources, Investigation, Formal analysis. **Yunjiao Zhang:** Resources, Investigation. **Jie Gu:** Resources, Formal analysis, Data curation. **Rongwei Zhou:** Visualization, Investigation. **Zhong Guo:** Writing – original draft, Supervision, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

We thank all members in Zhong Guo group for discussion.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.heliyon.2024.e33928.

References

[1] C.S. Dela Cruz, L.T. Tanoue, R.A. Matthay, Lung cancer: epidemiology, etiology, and prevention, Clin. Chest Med. 32 (4) (2011) 605–644.
[2] J.E. Larsen, J.D. Minna, Molecular biology of lung cancer: clinical implications, Clin. Chest Med. 32 (4) (2011) 703–740.
[3] J.R. Molina, et al., Non-small cell lung cancer: epidemiology, risk factors, treatment, and survivorship, Mayo Clin. Proc. 83 (5) (2008) 584–594.
[4] O. Rodak, et al., Current landscape of non-small cell lung cancer: epidemiology, histological classification, targeted therapies, and immunotherapy, Cancers 13 (18) (2021).
[5] Z. Chen, et al., Non-small-cell lung cancers: a heterogeneous set of diseases, Nat. Rev. Cancer 14 (8) (2014) 535–546.
[6] C. Genova, The long run towards personalized therapy in non-small-cell lung cancer: current state and future directions, Int. J. Mol. Sci. 24 (9) (2023).
[7] P. Krzyszczyk, et al., The growing role of precision and personalized medicine for cancer treatment, Technology (Singap World Sci) 6 (3–4) (2018) 79–100.
[8] J.C. Restrepo, et al., Advances in genomic data and biomarkers: revolutionizing NSCLC diagnosis and treatment, Cancers 15 (13) (2023) 3474.
[9] A. Bodaghi, N. Fattahi, A. Ramazani, Biomarkers: promising and valuable tools towards diagnosis, prognosis and treatment of Covid-19 and other diseases, Heliyon 9 (2) (2023) e13323.
[10] K.D. Davis, et al., Discovery and validation of biomarkers to aid the development of safe and effective pain therapeutics: challenges and opportunities, Nat. Rev. Neurol. 16 (7) (2020) 381–400.
[11] S.J. Klempner, et al., Tumor mutational burden as a predictive biomarker for response to immune checkpoint inhibitors: a review of current evidence, Oncol. 25 (1) (2020) e147–e159.
[12] M. Yarchoan, et al., PD-L1 expression and tumor mutational burden are independent biomarkers in most cancers, JCI Insight 4 (6) (2019).
[13] L.E. Hendriks, E. Rouleau, B. Besse, Clinical utility of tumor mutational burden in patients with non-small cell lung cancer treated with immunotherapy, Transl. Lung Cancer Res. 7 (6) (2018) 647–660.
[14] R. Baghban, et al., Tumor microenvironment complexity and therapeutic implications at a glance, Cell Commun. Signal. 18 (1) (2020) 59.
[15] E. Henke, R. Nandigama, S. Ergun, Extracellular matrix in the tumor microenvironment and its impact on cancer therapy, Front. Mol. Biosci. 6 (2019) 160.
[16] A. Scott, R. Salgia, Biomarkers in lung cancer: from early detection to novel therapeutics and decision making, Biomarkers Med. 2 (6) (2008) 577–586.
[17] D.R. Wang, X.L. Wu, Y.L. Sun, Therapeutic targets and biomarkers of tumor immunotherapy: response versus non-response, Signal Transduct. Targeted Ther. 7 (1) (2022) 331.
[18] V.K. Sarhadi, G. Armengol, Molecular biomarkers in cancer, Biomolecules 12 (8) (2022) 1021.
[19] E.R. Malone, et al., Molecular profiling for precision cancer therapies, Genome Med. 12 (1) (2020) 8.
[20] Y. Lei, et al., Progress and challenges of predictive biomarkers for immune checkpoint blockade, Front. Oncol. 11 (2021) 617335.
[21] R. Bai, et al., Predictive biomarkers for cancer immunotherapy with immune checkpoint inhibitors, Biomark. Res. 8 (2020) 34.
[22] T.A. Chan, et al., Development of tumor mutation burden as an immunotherapy biomarker: utility for the oncology clinic, Ann. Oncol. 30 (1) (2019) 44–56.
[23] M. Yi, et al., Biomarkers for predicting efficacy of PD-1/PD-L1 inhibitors, Mol. Cancer 17 (1) (2018) 129.
[24] M. Franklin, E. Connolly, T. Hussell, Recruited and tissue-resident natural killer cells in the lung during infection and cancer, Front. Immunol. 13 (2022) 887503.
[25] N.K. Wolf, D.U. Kissiov, D.H. Raulet, Roles of natural killer cells in immunity to cancer, and applications to immunotherapy, Nat. Rev. Immunol. 23 (2) (2023) 90–105.
[26] E.M. Levy, M.P. Roberti, J. Mordoh, Natural killer cells in human cancer: from biological functions to clinical applications, J. Biomed. Biotechnol. 2011 (2011) 676198.
[27] Y. Zeng, X. Lv, J. Du, Natural killer cell-based immunotherapy for lung cancer: challenges and perspectives, Oncol. Rep. 46 (5) (2021), 232.
[28] K.A. Shaver, T.J. Croom-Perez, A.J. Copik, Natural killer cells: the linchpin for successful cancer immunotherapy, Front. Immunol. 12 (2021) 679117.
[29] T. Bald, et al., The NK cell-cancer cycle: advances and new challenges in NK cell-based immunotherapies, Nat. Immunol. 21 (8) (2020) 835–847.
[30] N.K. Altorki, et al., The lung microenvironment: an important regulator of tumour growth and metastasis, Nat. Rev. Cancer 19 (1) (2019) 9–31.
[31] S.K. Larsen, Y. Gao, P.H. Basse, NK cells in the tumor microenvironment, Crit. Rev. Oncog. 19 (1–2) (2014) 91–105.
[32] I. Navin, M.T. Lam, R. Parihar, Design and implementation of NK cell-based immunotherapy to overcome the solid tumor microenvironment, Cancers 12 (12) (2020).
[33] S.Y. Wu, et al., Natural killer cells in cancer biology and therapy, Mol. Cancer 19 (1) (2020) 120.
[34] C. Zhang, Y. Hu, C. Shi, Targeting natural killer cells for tumor immunotherapy, Front. Immunol. 11 (2020) 60.
[35] M. Cheng, et al., NK cell-based immunotherapy for malignant diseases, Cell. Mol. Immunol. 10 (3) (2013) 230–252.

[36] C. Zalfa, S. Paust, Natural killer cell interactions with myeloid derived suppressor cells in the tumor microenvironment and implications for cancer immunotherapy, Front. Immunol. 12 (2021) 633205.
[37] N. Tumino, et al., Interaction between MDSC and NK cells in solid and hematological malignancies: impact on HSCT, Front. Immunol. 12 (2021) 638841.
[38] O. Melaiu, et al., Influence of the tumor microenvironment on NK cell function in solid tumors, Front. Immunol. 10 (2019) 3038.
[39] D. Lindau, et al., The immunosuppressive tumour network: myeloid-derived suppressor cells, regulatory T cells and natural killer T cells, Immunology 138 (2) (2013) 105–115.
[40] Y. Tie, et al., Immunosuppressive cells in cancer: mechanisms and potential therapeutic targets, J. Hematol. Oncol. 15 (1) (2022) 61.
[41] D. Huang, et al., Advances in single-cell RNA sequencing and its applications in cancer research, J. Hematol. Oncol. 16 (1) (2023) 98.
[42] D. Jovic, et al., Single-cell RNA sequencing technologies and applications: a brief overview, Clin. Transl. Med. 12 (3) (2022) e694.
[43] G. Chen, B. Ning, T. Shi, Single-cell RNA-seq technologies and related computational data analysis, Front. Genet. 10 (2019) 317.
[44] X.Y. Wen, et al., Integrating single-cell and bulk RNA sequencing to predict prognosis and immunotherapy response in prostate cancer, Sci. Rep. 13 (1) (2023) 15597.
[45] P. Song, et al., Identification and validation of a novel signature based on NK cell marker genes to predict prognosis and immunotherapy response in lung adenocarcinoma by integrated analysis of single-cell and bulk RNA-sequencing, Front. Immunol. 13 (2022) 850745.
[46] A. Haque, et al., A practical guide to single-cell RNA-sequencing for biomedical research and clinical applications, Genome Med. 9 (1) (2017) 75.
[47] K.J. Bienkowska, C.J. Hanley, Improved understanding of NSCLC immunotherapy response mechanisms from single-cell analysis, Transl. Lung Cancer Res. 12 (8) (2023) 1807–1811.
[48] Q. Xue, et al., Promising immunotherapeutic targets in lung cancer based on single-cell RNA sequencing, Front. Immunol. 14 (2023) 1148061.
[49] K.H. Prazanowska, S.B. Lim, An integrated single-cell transcriptomic dataset for non-small cell lung cancer, Sci. Data 10 (1) (2023) 167.
[50] Q. Li, et al., Molecular profiling of human non-small cell lung cancer by single-cell RNA-seq, Genome Med. 14 (1) (2022) 87.
[51] Y. Wang, et al., Multidirectional characterization of cellular composition and spatial architecture in human multiple primary lung cancers, Cell Death Dis. 14 (7) (2023) 462.
[52] C. Dominguez Conde, et al., Cross-tissue immune cell analysis reveals tissue-specific features in humans, Science 376 (6594) (2022) eabl5197.
[53] F.A. Wolf, P. Angerer, F.J. Theis, SCANPY: large-scale single-cell gene expression data analysis, Genome Biol. 19 (1) (2018) 15.
[54] F. Tang, et al., A pan-cancer single-cell panorama of human natural killer cells, Cell 186 (19) (2023) 4235–4251 e20.
[55] S. Mariathasan, et al., TGFbeta attenuates tumour response to PD-L1 blockade by contributing to exclusion of T cells, Nature 554 (7693) (2018) 544–548.
[56] P. Jiang, et al., Signatures of T cell dysfunction and exclusion predict cancer immunotherapy response, Nat. Med. 24 (10) (2018) 1550–1558.
[57] P. Shafer, L.M. Kelly, V. Hoyos, Cancer therapy with TCR-engineered T cells: current strategies, challenges, and prospects, Front. Immunol. 13 (2022) 835762.
[58] V. Thorsson, et al., The immune landscape of cancer, Immunity 48 (4) (2018) 812–830 e14.
[59] M. Wu, D. Kong, Y. Zhang, SPON2 promotes the bone metastasis of lung adenocarcinoma via activation of the NF-kappaB signaling pathway, Bone 167 (2023) 116630.
[60] Z. Chen, et al., Tumor-derived exosomal HOTAIRM1 regulates SPON2 in CAFs to promote progression of lung adenocarcinoma, Discov Oncol 13 (1) (2022) 92.
[61] K. Peng, X. Ren, Q. Ren, NcRNA-mediated upregulation of CAMK2N1 is associated with poor prognosis and tumor immune infiltration of gastric cancer, Front. Genet. 13 (2022) 888672.
[62] H.M. Koh, D.H. Song, Prognostic role of Rab27A and Rab27B expression in patients with non-small cell lung carcinoma, Thorac Cancer 10 (2) (2019) 143–149.
[63] D.P. Wang, et al., CtBP2 promotes proliferation and reduces drug sensitivity in non-small cell lung cancer via the Wnt/beta-catenin pathway, Neoplasma 65 (6) (2018) 888–897.
[64] C.C. Fan, et al., EFHD2 contributes to non-small cell lung cancer cisplatin resistance by the activation of NOX4-ROS-ABCC1 axis, Redox Biol. 34 (2020) 101571.
[65] C.C. Fan, et al., EFHD2 promotes epithelial-to-mesenchymal transition and correlates with postsurgical recurrence of stage I lung adenocarcinoma, Sci. Rep. 7 (1) (2017) 14617.
[66] X. Liu, L. Chen, T. Zhang, Increased GOLM1 expression independently predicts unfavorable overall survival and recurrence-free survival in lung adenocarcinoma, Cancer Control 25 (1) (2018) 1073274818778001.
[67] Y. Meng, et al., Clinical prognostic value of the PLOD gene family in lung adenocarcinoma, Front. Mol. Biosci. 8 (2021) 770729.
[68] P. Martinez-Camblor, J.C. Pardo-Fernandez, Parametric estimates for the receiver operating characteristic curve generalization for non-monotone relationships, Stat. Methods Med. Res. 28 (7) (2019) 2032–2048.
[69] A.C. Bretz, et al., Domatinostat favors the immunotherapy response by modulating the tumor immune microenvironment (TIME), J Immunother Cancer 7 (1) (2019) 294.