

# Immune subgroup analysis for non-small cell lung cancer may be a good choice for evaluating therapeutic efficacy and prognosis

Yuan Tian<sup>1,2,3,\*</sup>, Jingnan Wang<sup>1,\*</sup>, Qing Wen<sup>4</sup>, Guohai Su<sup>5</sup>, Yuping Sun<sup>6,1</sup>

<sup>1</sup>Department of Oncology, Jinan Central Hospital, Shandong University, Jinan 250013, Shandong, P.R. China

<sup>2</sup>Department of Radiotherapy Oncology, Shandong Provincial Qianfoshan Hospital, Shandong University, Jinan 250014, Shandong, P.R. China

<sup>3</sup>Department of Radiotherapy Oncology, Shandong Provincial Qianfoshan Hospital, The First Hospital Affiliated with Shandong First Medical University, Jinan 250014, Shandong, P.R. China

<sup>4</sup>Jinan Clinical Research Center of Shandong First Medical University, Jinan 250013, Shandong, P.R. China

<sup>5</sup>Department of Cardiovascular Diseases, Jinan Central Hospital Affiliated to Shandong University, Jinan 250013, Shandong, P.R. China

<sup>6</sup>Department of Oncology, Jinan Central Hospital Affiliated to Shandong First Medical University, Jinan 250013, Shandong, P.R. China

\*Equal contribution

**Correspondence to:** Guohai Su, Yuping Sun; **email:** [sg1630@zxyy.cn](mailto:sg1630@zxyy.cn), [199057020185@email.sdu.edu.cn](mailto:199057020185@email.sdu.edu.cn)

**Keywords:** immune subgroup, NSCLC, therapeutic efficacy, prognosis

**Received:** December 2, 2020

**Accepted:** March 4, 2021

**Published:** May 11, 2021

**Copyright:** © 2021 Tian et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/3.0/) (CC BY 3.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

## ABSTRACT

Due to its effectiveness, cancer immunotherapy has attracted widespread attention from clinicians and scientific researchers. Numerous studies have proven that effective stratification of cancer patients would promote the personalized application of immunotherapy. Therefore, we used the transcriptome data of nearly 1,000 patients with non-small cell lung cancer (NSCLC) to construct a new immune subgroup. We found that the new immune subgroup, named cluster 2, was a mixture of lung adenocarcinoma (LUAD) and lung squamous cell carcinoma (LUSC), and showed poor overall survival, which was further verified in the independent validation set. Immune infiltration correlation analysis showed that the Mast cell type and its status subdivisions had a predictive effect on the prognosis of NSCLC, especially in LUAD. Phenotypic analysis suggested that epithelial-mesenchymal transition (EMT) was positively correlated with immunosuppression, supporting the correlation between tumor phenotype and immune background. Although immune subtypes failed to significantly distinguish the progression-free survival (PFS) of immunotherapy patients, they showed the expected trend; the sample size needs to be further expanded for verification. In addition, some results indicated that the two cancer types, LUAD and LUSC, might require independent analyses.

## INTRODUCTION

Tumor immunotherapy has developed rapidly in recent years. Due to its effectiveness and safety [1, 2], it has attracted widespread attention from researchers and

clinicians, especially immunotherapy related to PD-1/PD-L1 inhibitors for lung cancers [3–15]. However, even in patients with the same type of tumor, immunotherapy response rates were quite variable, and in some patients the rates were relatively low [3–15].

According to the corresponding immune indicators, researchers have reported that the stratification of cancer patients was helpful for predicting the patients' prognosis or immunotherapy response rates [16, 17]. For clear cell renal cell carcinoma, it was reported that the potential interaction between immune phenotype and somatic changes affected the efficacy of PD-1 inhibitors [18, 19]. The correlation between immune subgroup and immunotherapy response was also observed in a study on non-small cell lung cancer (NSCLC) [20]. Based on the aforementioned research [16–20], it was speculated that there may be some inherent correlations between immune subgroup and NSCLC, which affect the response rate of immunotherapy and may further affect the prognosis of patients.

With the development of sequencing technology and bioinformatic analysis methods, a large quantity of cancer-related genome, transcriptome, and immune-related information can be obtained through sequencing or reanalysis of sequencing results [21–24]. Bioinformatic analysis has become a common approach for investigating cancer-related indicators [21–24]. It had been reported that the intermediate immune infiltration cluster (Cluster B) was associated with a worse prognosis independently of known clinicopathological features in breast cancers [25]. In order to clarify the inherent relationship between immune subgroup and NSCLC, bioinformatic analyses and online sequencing results were used to construct a new kind of immune subgroup based on the immune infiltration cluster method [25]. Bioinformatic analyses were completed mainly using the R programming language.

## **MATERIALS AND METHODS**

### **The cancer genome atlas (TCGA) expression profile data**

TCGA Pan-Cancer Atlas integrated pan-cancer RNA-seq data was downloaded from the Genomic Data Commons (GDC) website (<https://gdc.cancer.gov/about-data/publications/pancanatlas>). Based on the disease and sample labels, only samples from the original site of LUAD (n=506) and LUSC (n=482) (\*-01A) were extracted. Genes for which the expression was 0 in all samples were removed, and fragments per kilobase of exon model per million mapped reads (FPKM) values were normalized to log2. Only one technical duplication remained.

### **Gene expression omnibus (GEO) validation data download and preprocessing**

GSE4573 chip data was downloaded from the GEO database, and the tumor samples (HG-U133A platform)

were extracted, comprising a total of 130 samples. The chip data was standardized based on the Robust Multichip Average method. The expression value of each gene was calculated based on the relationship between the corresponding probe and the gene. For cases in which a gene corresponded to multiple probes, we chose the mean value of these probes as the expression value of the gene. The sample characteristics of TCGA and GEO are shown in Table 1.

### **Identification of immune subtypes based on consensus clustering**

770 immune-related genes were collected from the nCounter® PanCancer Immune Profiling Panel, and the correlation matrix was calculated among samples based on their expression. Consensus clustering was implemented based on the R package ConsensusClusterPlus, and the clustering method and distance measurement corresponded to Ward.D2 and Pearson correlation distance, respectively. Based on the change in value of the area under the consistent connected dominating set (CDS) and CDS curve, the number of consistent clusters was determined to be 4. Since there was only one sample in category 4, it was removed in the subsequent analysis.

### **Immune cell infiltration ratio and immune-related evaluation score**

The infiltration ratio of 22 immune cell types in tumor samples was calculated based on the gene expression deconvolution algorithm CIBERSORT [26]. The expression feature Leukocyte signature Matrix 22 (LM22) of immune cell signature genes was constructed using the CIBERSORT algorithm in advance. This expression matrix contained the expression patterns of 547 genes in 22 types of cells [naive B cells, memory B cells, Plasma cells, CD8 T cells, naive CD4 T cells, resting memory CD4 T cells, activated memory CD4 T cells, follicular helper T cells, regulatory T cells (Tregs), gamma delta T cells, resting NK cells, activated NK cells, Monocytes, M0 Macrophages, M1 Macrophages, M2 Macrophages, resting Dendritic cells, activated Dendritic cells, resting Mast cells, activated Mast cells, Eosinophils, and Neutrophils]. For the normalized expression profile data, we used the default immune cell feature matrix LM22, randomly 100 times to obtain the immune infiltration ratio of each sample. Then, the state of the immune cells was distinguished according to the method used by Thorsson [27]. When calculating and analyzing the differences in the proportion of immune cell infiltration, the proportion of infiltration for each cell type was first standardized by z-score; then, the significance of the difference was calculated based on the Kruskal-Wallis test. Lymphoid

**Table 1. Basic characteristics of the data included in the study.**

	TCGA_NSCLC (N=996)	GSE4573 (N=130)	Overall (N=1126)
<b>Histology</b>			
LUAD	509 (51.1%)	0 (0%)	509 (45.2%)
LUSC	487 (48.9%)	130 (100%)	617 (54.8%)
<b>Stage</b>			
Stage I	509 (51.1%)	73 (56.2%)	582 (51.7%)
Stage II	277 (27.8%)	34 (26.2%)	311 (27.6%)
Stage III	165 (16.6%)	23 (17.7%)	188 (16.7%)
Stage IV	33 (3.3%)	0 (0%)	33 (2.9%)
Missing	12 (1.2%)	0 (0%)	12 (1.1%)
<b>Gender</b>			
Female	401 (40.3%)	48 (36.9%)	449 (39.9%)
Male	595 (59.7%)	82 (63.1%)	677 (60.1%)
<b>Age (years)</b>			
Mean (SD)	66.2 (9.34)	67.5 (9.86)	66.4 (9.41)
Median [Min, Max]	67.0 [38.0, 90.0]	68.0 [42.0, 91.0]	67.0 [38.0, 91.0]
Missing	28 (2.8%)	0 (0%)	28 (2.5%)
<b>Smoking_Pack_years</b>			
Mean (SD)	47.5 (28.9)	62.7 (43.9)	49.5 (31.6)
Median [Min, Max]	40.0 [0.150, 200]	60.0 [0, 300]	45.0 [0, 300]
Missing	237 (23.8%)	18 (13.8%)	255 (22.6%)

and myeloid scores were calculated based on Nanodissect [28].

### Prediction of immune subgroups based on binomial logistic regression

An immune subgroup prediction model was developed based on the binomial logistic regression method, and optimized by L1 regularization (lasso) and full subset regression. The model was validated by shuffle-split with five-fold cross-validation. These five models were integrated into an ensemble model to output the final classification index.

### Survival analysis

Kaplan-Meier graphs were generated using the R package survminer. The Maximum Selection Test (Maximally Selected Log-Rank Statistic, R package survminer::surv\_cutpoint) method was used to identify the best cutpoint for high/low infiltration or high/low expression.

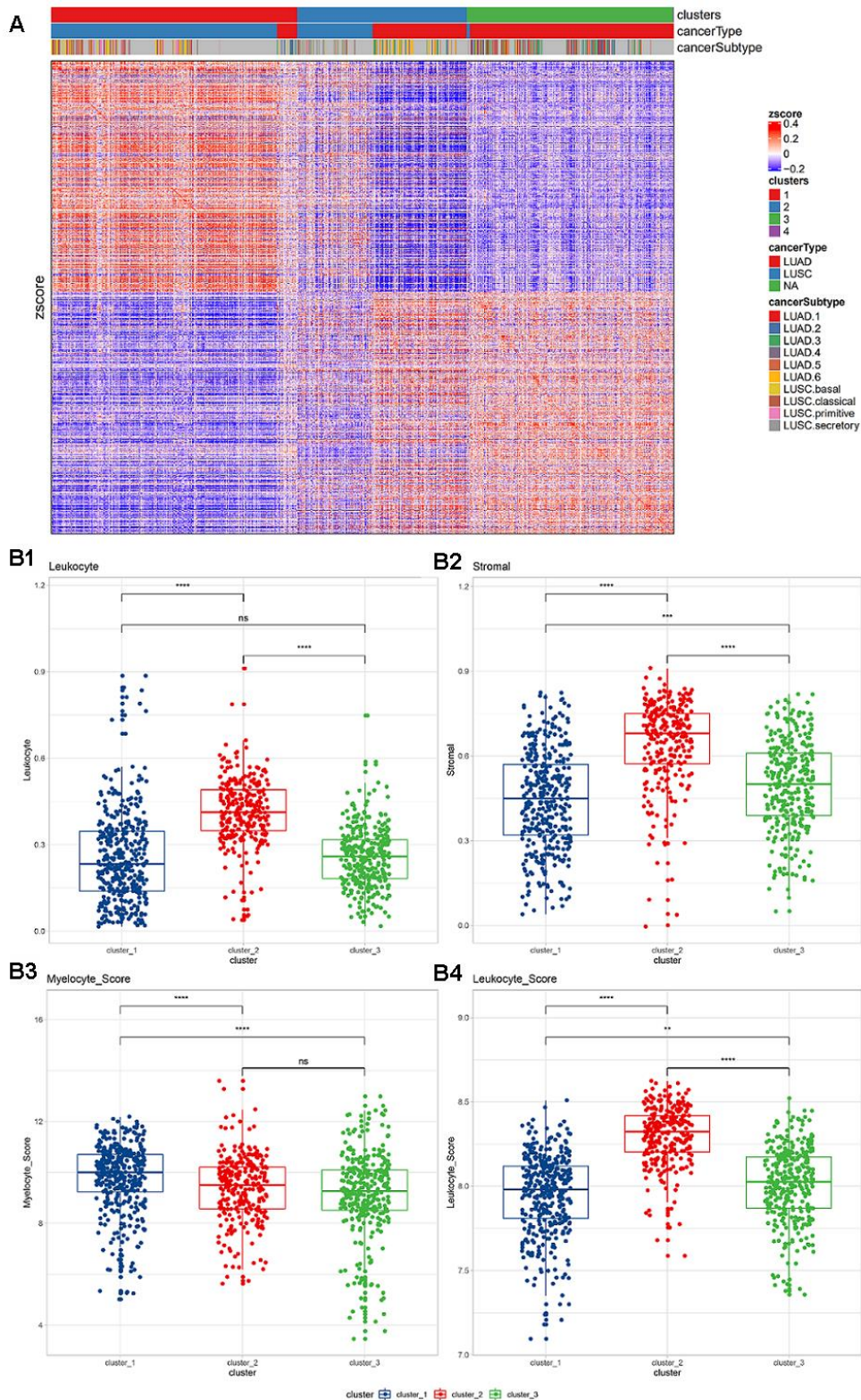
## RESULTS

### Constructing a new immune subgroup of NSCLC

The flow diagram is provided in Supplementary Figure 1. Based on the expression levels of 770 immune-related

genes derived from the nCounter® Pan-Cancer Immune Profiling array, a consistent clustering of 988 NSCLC samples was put into practice. Based on the change in the area under the consistent CDS and CDS curve, the number of consistent clusters was selected to be 4 (Supplementary Figure 2). Since there was only one sample in category 4, it was removed in the subsequent analysis. In the identified subgroups, most of the samples in cluster 1 (n=394) were squamous carcinomas, almost all of the samples in cluster 3 (n=331) were adenocarcinoma, and the samples in cluster 2 (n=272) were a mixture of the two cancer types (Figure 1A). To confirm that these three clusters were related to the tumor immune microenvironment (TIME), the Nanodissect algorithm was used to score total lymphocytes and myeloid cell infiltration. The three clusters were found to be significantly correlated with the scores of lymphocytes, myeloid cells and stromal cells (Wilcoxon rank-sum test,  $p < 0.0001$ , Figure 1B).

Based on pathological evaluation, the lymphatic invasion of cluster 2 was the highest and differed significantly from the other two groups, while no difference in lymphatic invasion was found between clusters 1 and 3. The levels of stromal cells in the three subgroups were found to be significantly different; cluster 2 was considered to have the highest level (Figure 1B1, 1B2). Based on the expression level of



**Figure 1.** (A) Expression of immune-related genes in TCGA non-small cell lung cancer data. Pearson correlation distance and ward.D2 were used for unsupervised clustering. The top annotations represent the subgroups, cancer types, and TCGA cancer subtypes identified based on consistent clustering. (B) Lymphocyte and myeloid cell infiltration and stromal cell distribution among subgroups. (B1) Lymphocyte infiltration in different clusters; (B2) stromal cell distribution in different clusters; (B3) myelocyte infiltration score in different clusters; (B4) lymphocyte infiltration score in different clusters. The Wilcoxon rank-sum test (\*,  $P < 0.05$ ; \*\*,  $P < 0.01$ ; \*\*\*,  $P < 0.001$ ; \*\*\*\*,  $P < 0.0001$ ) was used for statistical difference testing among different groups.



infiltration score, cluster 2 was also shown with the highest level of lymphatic invasion, while myeloid infiltration was found to be gradually increased in the three subgroups (Figure 1B3, 1B4).

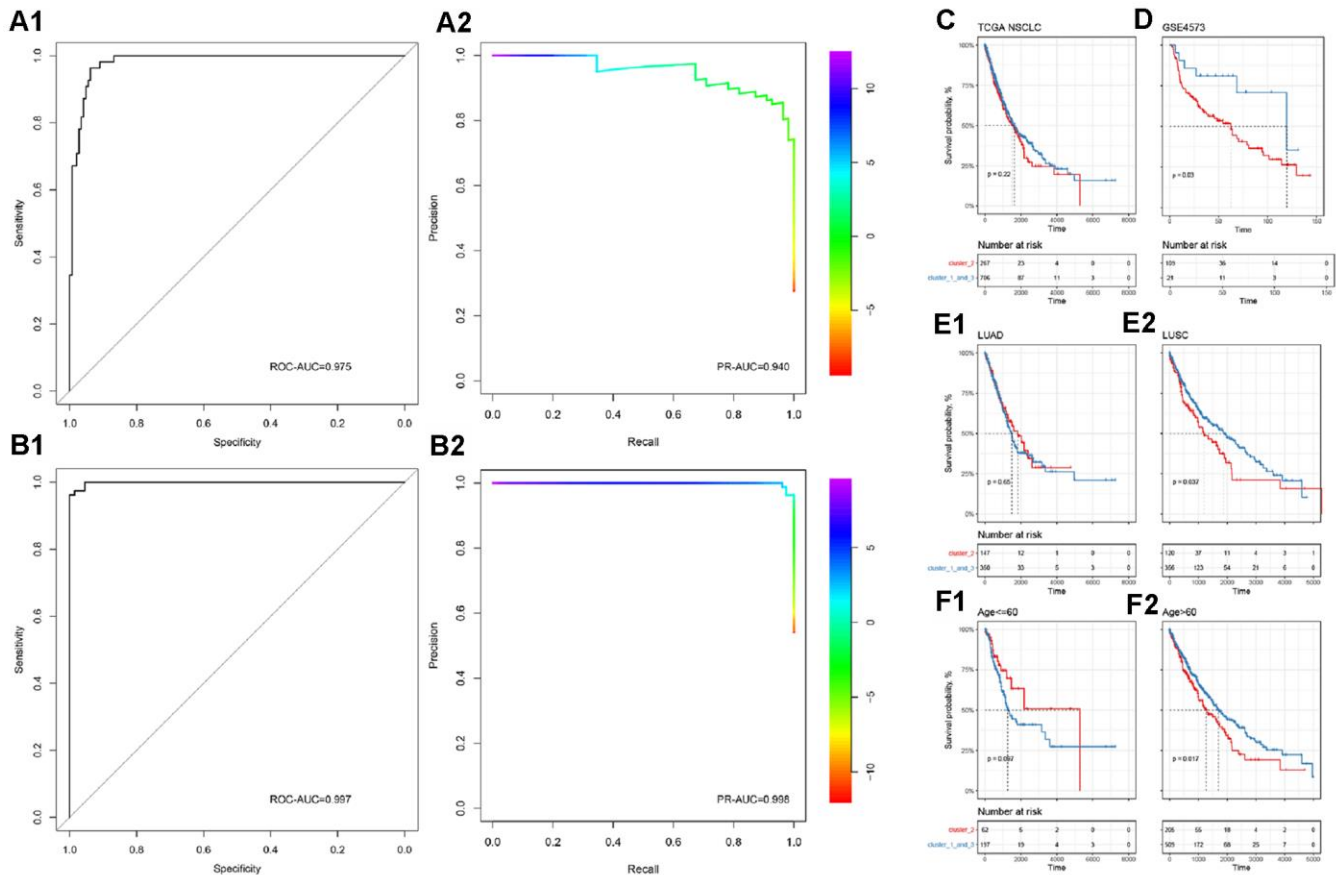
Among the three clusters, the enrichment scores of all immune-related pathways showed significant differences. Most of the pathways showed stronger signal in cluster 2, and were significantly higher than the other two groups (Kruskal test, \*,  $P < 0.05$ ; \*\*,  $P < 0.01$ ; \*\*\*,  $P < 0.001$ ; \*\*\*\*,  $P < 0.0001$ ; Supplementary Figure 3). Among the three clusters, all suppressive immune checkpoint genes showed significant expression differences. The vast majority of pathways showed significantly high expression in cluster 2 (Kruskal test, \*,  $P < 0.05$ ; Supplementary Figure 4). Among them, the macrophage CD86 and the widely used ICB target-CD274 are both significantly highly expressed in cluster 2.

## Building an immune subgroup prediction model

In order to accurately distinguish lung cancer subgroups without relying on unsupervised clustering, we then used binomial logistic regression and L1 regularization (lasso) to construct a subgroup classifier to distinguish cluster 2 from other subgroups (cluster 1 and cluster 3). It was confirmed by five-fold crossover that our model had good predictive power (ROC-AUC=0.975, PR-AUC=0.940, Figure 2A, 2B). Furthermore, we performed a new round of binomial logistic regression to distinguish cluster 1 and cluster 3, which also achieved extremely high performance (ROC-AUC=0.997, PR-AUC=0.998).

## Prognostic efficacy of immune subgroups

To complete our observations, we mapped patient survival based on the identified immune subgroups.



**Figure 2. Subgroup prediction model based on binomial logistic regression and prognostic efficacy of immune subgroups.** (A) Distinguishing the predictive power of cluster 2 from the other two groups. (B) Distinguishing the predictive power of cluster 1 from cluster 3. The upper part is ROC-AUC, and the lower part is PR-AUC. (C) In all NSCLC samples, immune subgroups cannot significantly distinguish the overall survival of patients. (D) In the GEO independent verification set, the immune subgroup can significantly distinguish the prognosis of patients. Among patients with lung squamous cell carcinoma (E) or those older than 60 years (F), immune subgroups can significantly distinguish the prognosis of the patients.

Unfortunately, immune subgroup failed to distinguish the overall survival of patients effectively in TCGA-NSCLC (log-rank test,  $p > 0.05$ , Figure 2C). In the independent validation set, patient survival could be distinguished effectively by the identified immune subgroups (log-rank test,  $p < 0.05$ , Figure 2D). After considering the effects of age, tumor stage, gender, and cancer type, it was found that the current immune subgroup could significantly distinguish the overall survival from others when patients were over 60 years or had squamous cell carcinoma (log-rank test,  $p < 0.05$ , Figure 2E, 2F). In addition, progression-free survival (PFS) of Stage II NSCLC patients could also be significantly distinguished by immune subgroup (log-rank test,  $p < 0.05$ , Supplementary Figure 5).

### Molecular indicators of immune subgroups

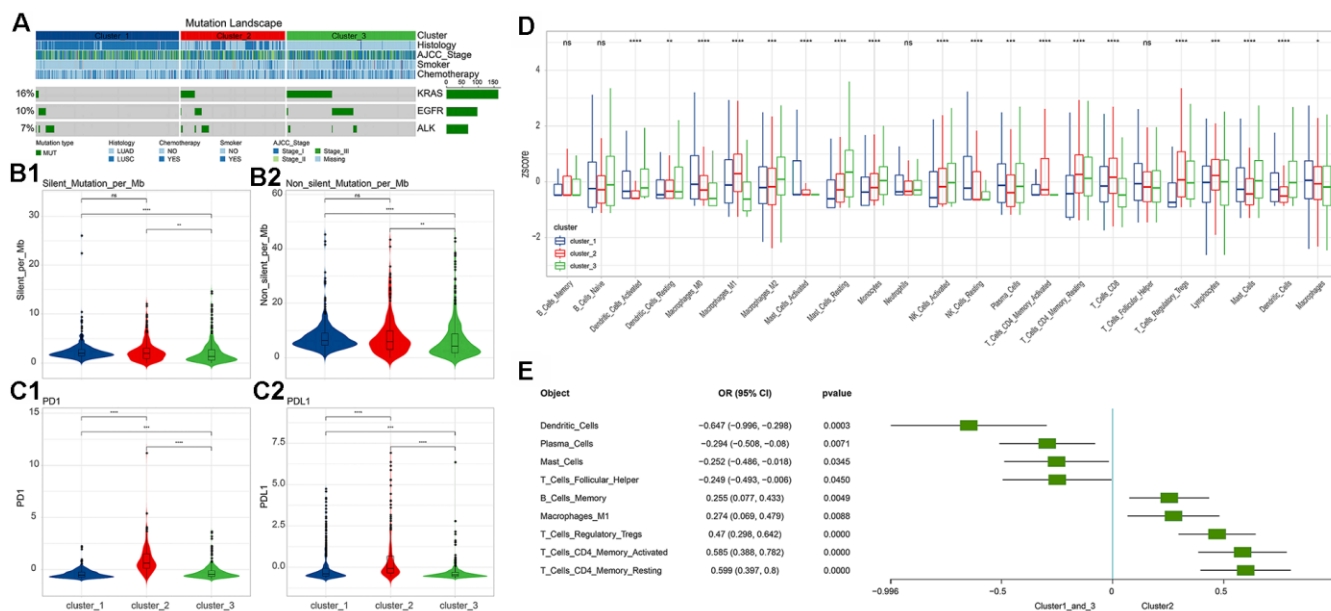
Subsequently, we systematically analyzed the molecular and clinical indicators of immune subgroups, including genomic alterations, pathological typing, and immune infiltration. By comparing the three immune subgroups, we found that EGFR and KRAS were significantly different in the subgroups (chi square test,  $p < 0.05$ , Figure 3A), and mainly appeared in cluster 3. EGFR and KRAS were mutually exclusive in cluster 2 and cluster 3 (chi square test,  $p < 0.05$ ). Analyzing the

correlation between the three subgroups and clinical information, we found that smoking was significantly related to immune subgroup, mainly apparent in cluster 3 (chi square test,  $p < 0.05$ ). However, the immune subgroup had nothing to do with American Joint Committee on Cancer (AJCC) Stage or chemotherapy (chi square test,  $p > 0.05$ ). In addition, the silent mutation and non-silent mutation loads of cluster 3 were found to be at the lowest level among the clusters, significantly lower than those of cluster 1 and cluster 2 (Wilcoxon rank-sum test, Figure 3B); there was no significant difference in tumor mutation burden (TMB) between cluster 2 and cluster 1. Turning our attention to PD-1 and PD-L1, we found that the expression levels of these two molecules (z-score of  $\log_2$ -transformed FPKM) in the cluster 2 subgroup were significantly higher than those of the other two subgroups (Wilcoxon rank-sum test, Figure 3C).

### Immune cell infiltration in different subgroups

Based on the CIBERSORT method, we calculated the immune cell infiltration of each sample.

We found that 22 immune cell types and their status subdivisions showed significant differences in the three subgroups. Among them, M1 Macrophages were found



**Figure 3. Differences in molecular indicators among immune subgroups and subtype-related immune cell infiltration.** (A) Mutation landscape in all clusters. KRAS and EGFR mutations occurred significantly in cluster 3 (chi square test,  $p < 0.05$ ). (B) Tumor mutation burden (Silent and Non-silent) in different clusters. Cluster 3 had the lowest tumor mutation burden. (C) Expression levels of PD-1/PD-L1 in different clusters. Cluster 2 had the highest levels of PD-1 and PD-L1 expression. The numerical statistical test was based on the Wilcoxon rank sum test (\*,  $P < 0.05$ ; \*\*,  $P < 0.01$ ; \*\*\*,  $P < 0.001$ ; \*\*\*\*,  $P < 0.0001$ ). (D) Differences in immune cell infiltration subtypes. Statistical significance was calculated using the Kruskal-Wallis test (\*,  $P < 0.05$ ; \*\*,  $P < 0.01$ ; \*\*\*,  $P < 0.001$ ; \*\*\*\*,  $P < 0.0001$ ). (E) Immune cell infiltration significantly associated with cluster 2 (based on logistics regression identification,  $p < 0.05$ ).

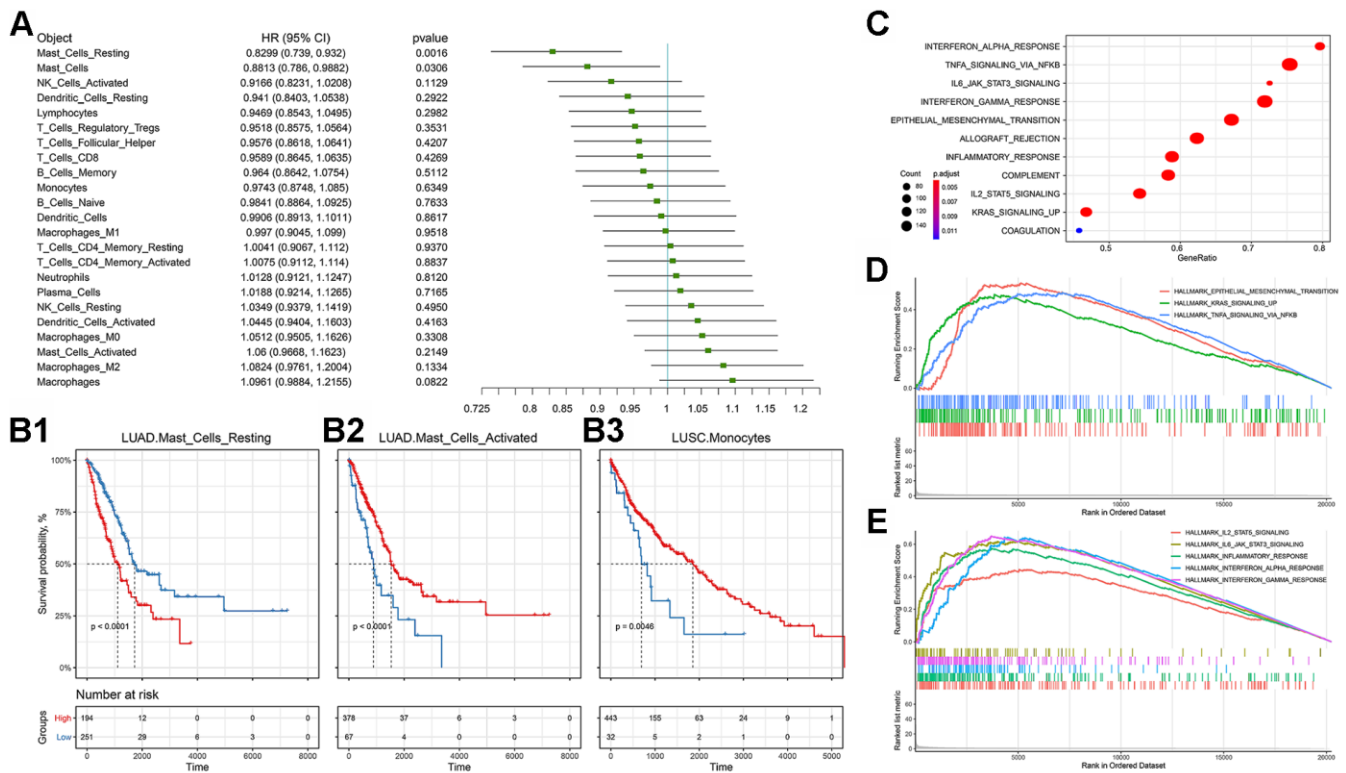
to be have the greatest infiltration in cluster 2, while resting Mast cells had the greatest infiltration in cluster 3 (Figure 3D and Supplementary Figure 6). Using the generalized linear model, we identified the immune cell types that could explain the difference between cluster 2 and cluster 1\_cluster 3 (Figure 3E).

We then observed the correlation between immune cell infiltration and the prognosis of NSCLC, and found that Mast cells and their status subdivisions were significantly associated with the prognosis of patients (Figure 4A). To further subdivide cancer types, we found that the Mast cell type and its status subdivisions could still distinguish the prognosis of patients in LUAD from the others, while Monocytes in LUSC patients was found to be related to the overall survival of patients (Figure 4B). Prior research suggested that the Mast cell type and its phenotype might be related to the prognosis of NSCLC [29]. Furthermore, it was reported that Mast cells might have a prognostic effect

on lung cancer [30]. Our results showed that activated and resting Mast cells were significantly associated with better and worse prognoses, respectively, which further supports the prognostic role of the Mast cell type and its status subdivisions in NSCLC. Unlike previous reports, in LUSC, we found that Monocytes were associated with a better prognosis [31–34].

### Functional differences in different subgroups

Based on the MSigDB cancer hallmark gene set, we used the Gene Set Enrichment Analysis (GSEA) method to identify the cluster 2-related functional differences. The results showed that compared with cluster\_1\_and\_3, cluster 2's dysregulated genes were enriched in a variety of pathways related to malignant cancer progression, including stem cell biology and EMT-related pathways corresponding to cancer-related functions (Figure 4C). Further visualizing the important cancer-related pathways, we found that cluster 2 had



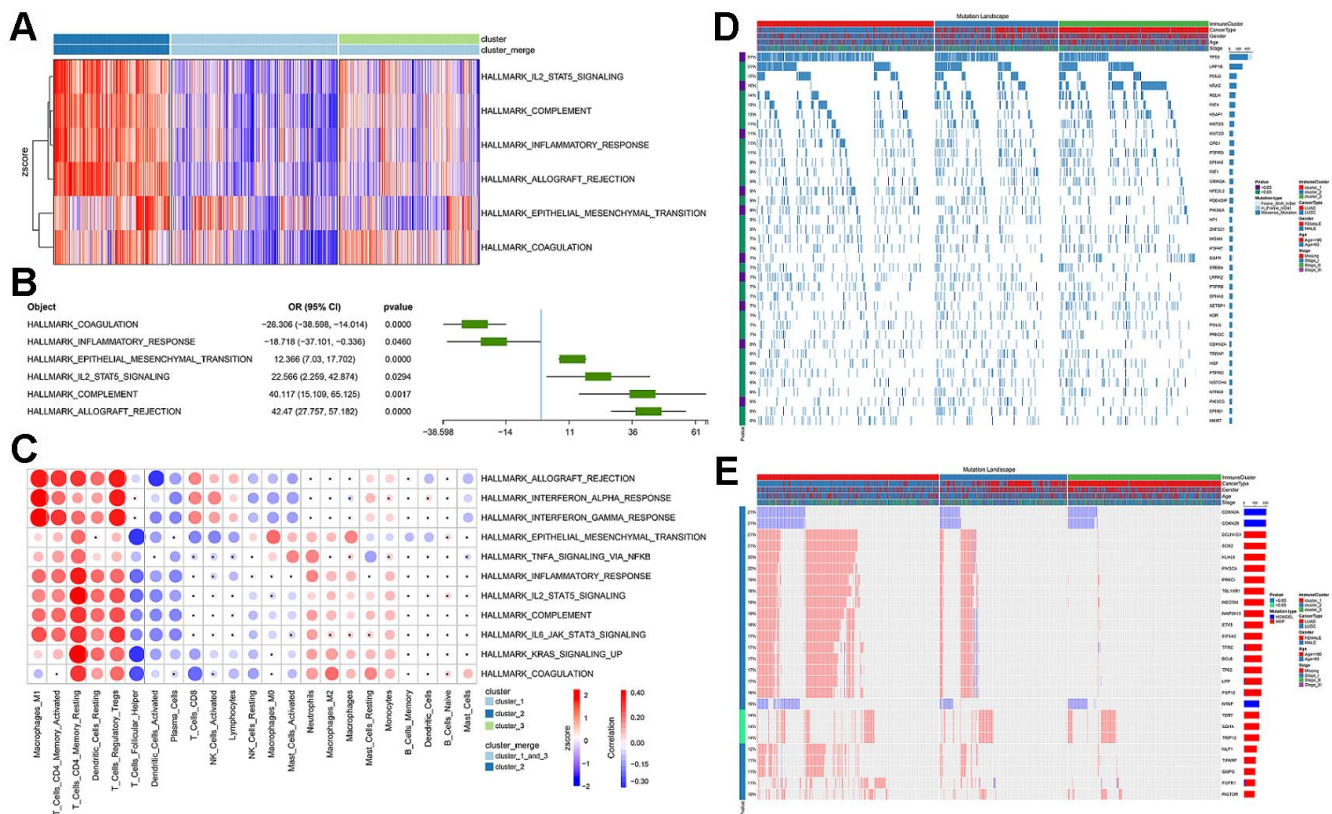
**Figure 4. Prognostic efficacy of immune cell infiltration in TCGA-NSCLC and the functional differences involved in cluster 2.** (A) Prognostic efficacy of immune cell infiltration in TCGA-NSCLC: Forest plot of the prognostic efficacy of immune cells. Single factor cox regression was used to evaluate the significance of the correlation between cell infiltration ratio and Overall Survival (OS). (B) Prognostic efficacy of immune cell infiltration in TCGA-NSCLC: In the corresponding metabolic subtypes, the Kaplan-Meier curve of the two groups of patients with high and low cell infiltration (Low). The log-rank test was used for calculating the difference in OS between the high and low groups. The division threshold was based on the maximum selection method. (C) Functional differences involved in cluster 2: Identifying the functional differences based on the GSEA method. The MSigDB cancer hallmark gene set was used. (D) Functional differences involved in cluster 2: Visualization of important cancer-related pathways. (E) Functional differences involved in cluster 2: Visualization of important immune-related pathways.

more EMT, KRAS and TNF signaling pathways, which was consistent with the worse prognosis associated with this cluster (Figure 4D, 4E). At the same time, we found that cluster 2 also had an enriched immune-related factor and active inflammatory response.

In order to further explain the relationship between immune subgroup and cancer cell phenotype, we calculated the signature score of the aforementioned gene set based on the Gene Set Variation Analysis (GSVA) method for each sample. This score reflected the activity of each pathway in the immune subgroup. Immune subgroups could be distinguished clearly by unsupervised clustering of gene set scores (Figure 5A). Through the unsupervised clustering of GSVA

enrichment scores, we found that those signatures appeared with high scores in cluster 2, but displayed erratic patterns in other subgroups. In order to formally determine which gene set score could explain cluster 2, we used the logistics model to test the contribution of each gene set to cluster 2 and cluster\_1\_and\_3. The results indicate that pathways such as EMT and IL2-STAT5-SIGNALING have a positive contribution to cluster 2, while cluster\_1\_and\_3 is related to inflammatory response (Figure 5B).

Since the immune subgroups were related to a variety of immune-related pathways, we then checked the relationship between immune cell infiltration and the aforementioned cancer-related pathways. We found that



**Figure 5. The functional differences involved in cluster 2 and the correlation between immune infiltration and cluster 2.**

(A) GSVA score spectrum of functional gene set in cluster 2. (B) Application of logistic regression to identify pathways that were significantly related to cluster 2. (C) Correlation between cluster function and immune infiltration. (D) Gene mutation characteristic blueprint of TCGA-NSCLC immune subgroup. 39 oncogenes mutated in >6% samples in the three immune subgroups of TCGA-NSCLC (missense mutation, dark blue; insert deletion, light blue). The bar graph on the right represents the mutation frequency of each gene in the sample population, and the colored bar graph on the left represents whether the occurrence of the gene mutation was related to the immune subgroup (chi-square test). Each column represents a patient, and the colored bar chart at the top indicates the immune subgroup to which the patient belonged: the subtype of NSCLC, the patient's gender, age group, and tumor grade. (E) Copy number variation blueprint of TCGA-NSCLC immune subgroups; the status of 26 oncogenes with copy number changes in >10% samples in the three immune subgroups of TCGA-NSCLC (copy number amplification, red; copy number deletion, blue). The bar graph on the right represents the copy number variation frequency of each gene in the sample population. The colored bar chart on the left represents whether the occurrence of copy number variation of the gene was related to the immune subgroup (chi-square test). Each column represents a patient, and the colored bar chart at the top indicates the immune subgroup to which the patient belonged: the subtype of NSCLC, the patient's gender, age group, and tumor grade.



the activity scores of these pathways were significantly correlated with infiltration of a variety of immune cells (Pearson's correlation test,  $p$  value $<0.05$ , Figure 5C). It was worth noting that EMT was significantly associated with all immune cells and their status subdivisions. The high EMT score was highly positively correlated with infiltration of a variety of inhibitory immune cells (e.g., M1 Macrophages, resting Dendritic cells, and regulatory T Cells), and negatively correlated with the infiltration of killer immune cells (e.g., CD8 T Cells and Activated NK Cells).

### **Subgroup-associated clinical features (EMT and proliferation) mutations, CNV, etc.**

We observed the mutation signature blueprints of TCGA-NSCLC samples in three different immune subtypes, and compared the mutations of some key oncogenes (from Oncology Knowledge Base; OncoKB) among different immune subgroups (Figure 5D). Most of the oncogenes with mutation frequency greater than 10% appeared with missense mutations. Among different immune subgroups, the mutations of TP53, KRAS, KMT2D, NFE2L2, PIK3CA, EGFR, LRRK2, SETBP1, CDKN2A, PIK3CG were found to be distributed in a significantly biased manner (Figure 5D) [35, 36]. In TCGA-NSCLC patients, we found that CDKN2A, CDKN2B and MTAP mainly had copy number deletions (Figure 5E), whereas PIK3CA, SOX2, BCL6 had a large number of copy number amplifications and were significantly concentrated in cluster 1.

### **Immunotherapy in different subgroups**

We further evaluated the relationship between immune subgroup and immunotherapy (Figure 6). Based on the previous model, we divided immunotherapy patients into cluster 2 and cluster\_1\_and\_3. We found that 31.4% (11/35) of PD patients, 22.9% (8/35) of PR patients, and 11.4% (4/35) of SD patients belonged to the cluster 2 subgroup, whereas 8.6% (3/35) of PD patients, 2.9% (1/35) of PR patients and 22.9% (8/35) of SD patients belong to the cluster\_1\_and\_3 subgroup. When comparing (CR & PR) vs. PD, there was no significant association between efficacy and immune subgroup (chi square test,  $p>0.05$ , Figure 6A). The PFS in the cluster 2 subgroup was shorter than that in the cluster\_1\_and\_3 subgroup, although no significant difference was found (log-rank test,  $p>0.05$ , Figure 6B). Observing the expression levels of PD-1, PD-L1, and PD-L2, we found that the expression levels of these genes in the cluster\_1\_and\_3 subgroup were significantly higher than those in the cluster 2 subgroup (Figure 6C).

## **DISCUSSION**

Among all patients with lung cancer, NSCLC accounts for more than 85% of cases, making it the most common pathological type in clinical work and the leading cause of cancer-related deaths [37]. In recent years, immunotherapy for NSCLC has been developing rapidly, especially for PD-1/PD-L1 inhibitors [38–42]. While tumor immunotherapy has achieved satisfactory anti-tumor efficacy, predictive indicators related to therapeutic efficacy and prognosis have also been gradually reported [16–19, 43]. Based on the aforementioned reports on immune predictive indicators, we designed the present study in order to comprehensively evaluate the impact of immune-related indicators or immune grouping methods on the therapeutic efficacy and prognosis of NSCLC [16–19, 43].

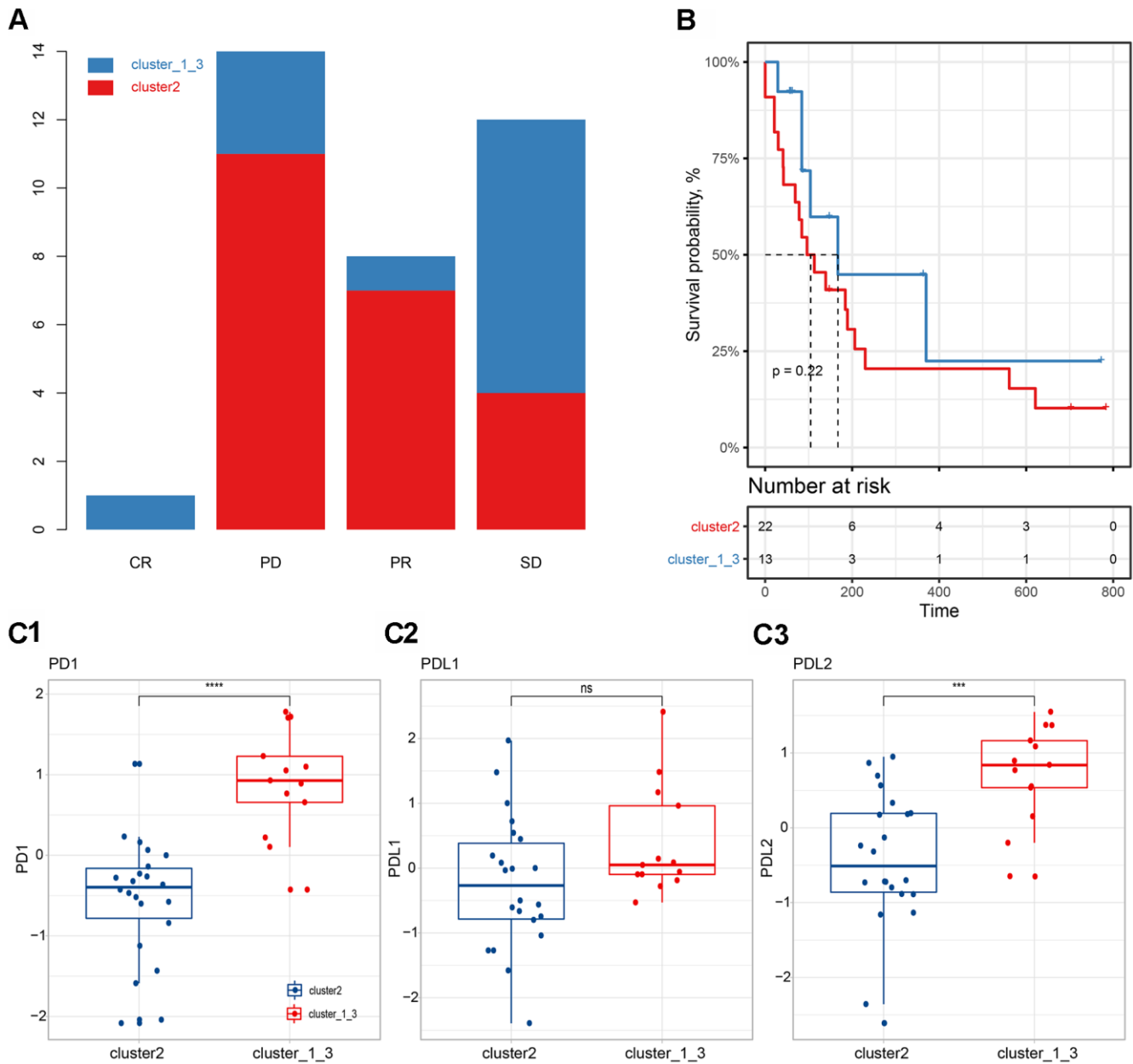
Based on the expression levels of 770 immune-related genes derived from the nCounter® PanCancer Immune Profiling array, 988 NSCLC samples were clustered uniformly, and three cluster subgroups were obtained (Figure 1A), cluster 1 ( $n=394$ ), cluster 2 ( $n=272$ ) and cluster 3 ( $n=331$ ), corresponding to the three most common pathological types in NSCLC: squamous cell carcinoma, adenosquamous cell carcinoma, and adenocarcinoma [44, 45]. We found that these three clusters were significantly correlated with the scores of lymphocytes, myeloid cells and stromal cells (Wilcoxon rank-sum test,  $p<0.0001$ , Figure 1B). Based on the Nanodissect algorithm to score total lymphocytes and myeloid cell infiltration (Figure 1B), we inferred that these three clusters may be related to the tumor immune microenvironment. This laid the foundation for us to further construct the immune subgroup prediction model (Figure 2A, 2B).

At present, immune checkpoint inhibitor therapies, especially for PD-1/PD-L1 inhibitors [3–15, 46, 47], have been developing rapidly for NSCLC, and a large number of indicators related to the efficacy of immunotherapy and patient prognosis have also been reported [48–50]. However, the prognostic efficacy of immune subgroups has rarely been reported. After considering the effects of age, tumor stage, gender, and cancer type, we found that overall survival could be significantly distinguished by immune subgroup when patients were over 60 years of age or had squamous cell carcinoma (log-rank test,  $p<0.05$ , Figure 2E, 2F). In addition, the PFS survival of Stage II NSCLC patients could also be significantly distinguished by immune subgroup (log-rank test,  $p<0.05$ , S Figure 3). In the cluster 2 subgroup, the expression levels of PD-1 and PD-L1 (z-scores of log<sub>2</sub>-transformed FPKM) were significantly higher than in the other two subgroups

(Wilcoxon rank-sum test, Figure 3C). Because the cluster 2 subgroup had higher TMB (Silent mutation per Mb and Non-silent mutation per Mb) and higher PD-1 and PD-L1 expressions, it may be implied to have a better response to immunotherapy [51, 52]. The aforementioned findings would be helpful for us to judge the prognosis of patients in clinical work.

Through the analysis of immune cell infiltration in different subgroups (Figure 3D, 3E), we found that the tumor-promoting immune cell types M1 Macrophages

[53], regulatory T Cells [54], activated memory CD4 T Cells [49, 55], and resting memory CD4 T Cells mainly existed in cluster 2, which might explain the reason for the worse prognosis with the cluster 2 subgroup. The analysis results showed that the Mast cell type and its status subdivisions are significantly associated with patient prognosis (Figure 4A). To further subdivide the cancer type, the Mast cell type and its status subdivisions could significantly distinguish the prognosis of patients in LUAD from others, while the presence of Monocytes in LUSC patients was



**Figure 6. Correlation between immune subgroups and the efficacy of immunotherapy.** (A) Efficacy of immunotherapy in different immune subgroups. (B) Progression-free survival in different immune subgroups. (C) Expression differences of PD-1 (C1), PD-L1 (C2) and PD-L2 (C3) in different subgroups.

significantly related to the overall survival of these patients (Figure 4B). Our analysis results were similar to those in prior reports [29, 30]. At the same time, the results also verified the feasibility of our analysis method from another aspect. Furthermore, our results showed that activated and resting Mast cells were significantly associated with better and worse prognoses, respectively, which further supports the prognostic role of the Mast cell type and its status subdivisions in NSCLC (Figure 4B1, 4B2). It has been reported that circulating Monocytes are associated with angiogenesis and poor prognosis in numerous cancers [31–34]. In LUSC, we found that Monocytes were associated with a better prognosis (Figure 4B3). This inconsistency may be due to the following reasons: (1) differences caused by the differentiation of monocytes into different cells (mainly including macrophages) after entering tissues; (2) specificity of tissue origin of LUSC.

Through our analysis, we found that more active pathways related to malignant progression of cancer were enriched in the cluster 2 subgroup (Figure 4C), which was consistent with the worse prognosis of this subgroup (Figure 4D). At the same time, we found that cluster 2 also has a strong immune-related factor and inflammatory response (Figure 4E). However, TCGA-NSCLC data are basically chemotherapy related. Therefore, the relationship between this strong immune response and the worse prognosis cannot be explained clearly.

In order to further explain the relationship between immune subgroups and cancer cell phenotype, we calculated the signature score of the aforementioned gene set based on the GSVA method for each sample. Immune subgroups could be distinguished clearly by unsupervised clustering of gene set scores (Figure 5A). The analysis results indicated that cluster 2 was positively influenced by pathways such as EMT and IL2-STAT5-SIGNALING, whereas cluster\_1\_and\_3 was related to inflammatory response (Figure 5B). Furthermore, we found that the activity scores of these pathways were significantly correlated with infiltration of a variety of immune cells (Figure 5C). Notably, EMT was significantly associated with all immune cells and their status subdivisions. EMT score was highly positively correlated with infiltration of a variety of inhibitory immune cells (e.g., M1 Macrophages, resting Dendritic cells, and regulatory T Cells; Figure 5C), and negatively correlated with the infiltration of killer immune cells (e.g., CD8 T Cells and activated NK cells; Figure 5C). This may further explain the poor prognosis of patients in the cluster 2 subgroup [56, 57].

We observed the mutation signature blueprints of TCGA-NSCLC samples in three different immune subtypes, and

compared the mutations of some key oncogenes (from OncoKB) among different immune subgroups (Figure 5D, 5E). The mutation frequency of TP53 was as high as 51%, which tended to appear in clusters 1 and 2, while cluster 3 had less mutation enrichment [58, 59]. As a rare oncogene related to the potential treatment of NSCLC, PIK3CA also showed a similar tendency for mutation enrichment, with more mutations in cluster 1 and 2 samples. It was reported that PIK3CA mutations could confer a relapse-free survival advantage for squamous cell carcinoma in NSCLC [35]. On the contrary, EGFR and KRAS mutations were abundantly enriched in cluster 3 samples, which was consistent with the function of suppressing immune infiltration as reported in some studies [36]. In our results, we also observed that the samples in cluster 3 had lower lymphocyte infiltration (Figure 1B). Both the enrichment of mutant genes and the alterations in copy number of related genes were basically consistent with previous reports [35, 36, 60, 61]. In other words, the use of our analysis method will make it easier for clinicians to judge the immune infiltration characteristics, mutations and prognostic survival of tumor patients, hence facilitating the formulation of treatment plans.

We further evaluated the relationship between immune subgroups and immunotherapy (Figure 6). The PFS in cluster 2 subgroups was shorter than those of cluster\_1\_and\_3 subgroups, although no significant difference was found (log-rank test,  $p > 0.05$ , Figure 6B). Observing the expression levels of PD-1, PD-L1, and PD-L2, we found that the expression levels of these genes in the cluster\_1\_and\_3 subgroup were significantly higher than those in the cluster 2 group (Figure 6C), which might partly explain the slightly better PFS of the cluster\_1\_and\_3 subgroup [51].

## CONCLUSIONS

The immune infiltration cluster 2 subgroup was a mixture of LUAD and LUSC, and showed poor overall survival, which was further verified in the independent validation set. Immune infiltration correlation analysis showed that the Mast cell type and its status subdivisions had a predictive effect on the prognosis of NSCLC, especially in LUAD.

### Statement of ethics

Since our study is based on online public database resources with no interaction with human subjects, no issues related to medical ethics are reported.

### Abbreviations

PD-1: Programmed Cell Death-1; PD-L1: Programmed Cell Death Ligand 1; NSCLC: Non-Small Cell Lung

Cancer; LUAD: Lung adenocarcinoma; LUSC: Lung squamous cell carcinoma; EMT: Epithelial mesenchymal transition; PFS: progression-free survival; TCGA: The Cancer Genome Atlas; GEO: Gene Expression Omnibus; CNV: Copy Number Variation; GDC: Genomic Data Commons; CDF: Cumulative Distribution Function; TIME: Tumor Immune Microenvironment; FPKM: Fragments Per Kilobase of Exon Model Per Million Mapped Reads; CDS: Connected Dominating Set; LM22: Leukocyte signature Matrix 22; AJCC: American Joint Committee on Cancer; TMB: Tumor Mutation Burden; GSVA: Gene Set Variation Analysis; GSEA: Gene Set Enrichment Analysis; OncoKB: Oncology Knowledge Base; PD: progressive disease; SD: stable disease; CR: complete response; PR: partial response.

## AUTHOR CONTRIBUTIONS

The corresponding authors (Yuping Sun and Guohai Su) had all the data and were responsible for the decision to submit it for publication. Yuan Tian and Jingnan Wang were responsible for data collection and analysis. Yuan Tian was appointed to write the draft of the manuscript. All the figures and tables were designed and typeset by Jingnan Wang. Qing Wen was responsible for all data verification in this article.

## CONFLICTS OF INTEREST

The authors declare that they have no conflicts of interest.

## FUNDING

This study was funded by Academic Promotion Program of Shandong First Medical University (2019QL025; Yuping Sun), Natural Science Foundation of Shandong Province (ZR2019MH042; Yuping Sun), and National Science and Technology Major Project of the Ministry of Science and Technology of China (2020ZX09201025; Guohai Su).

## REFERENCES

1. Kennedy LB, Salama AK. A review of cancer immunotherapy toxicity. *CA Cancer J Clin.* 2020; 70:86–104.  
<https://doi.org/10.3322/caac.21596>  
PMID:31944278
2. Anderson R, Theron AJ, Rapoport BL. Immunopathogenesis of Immune Checkpoint Inhibitor-Related Adverse Events: Roles of the Intestinal Microbiome and Th17 Cells. *Front Immunol.* 2019; 10:2254.  
<https://doi.org/10.3389/fimmu.2019.02254>  
PMID:31616428
3. Rittmeyer A, Barlesi F, Waterkamp D, Park K, Ciardiello F, von Pawel J, Gadgeel SM, Hida T, Kowalski DM, Dols MC, Cortinovis DL, Leach J, Polikoff J, et al, and OAK Study Group. Atezolizumab versus docetaxel in patients with previously treated non-small-cell lung cancer (OAK): a phase 3, open-label, multicentre randomised controlled trial. *Lancet.* 2017; 389:255–65.  
[https://doi.org/10.1016/S0140-6736\(16\)32517-X](https://doi.org/10.1016/S0140-6736(16)32517-X)  
PMID:27979383
4. Langer CJ, Gadgeel SM, Borghaei H, Papadimitrakopoulou VA, Patnaik A, Powell SF, Gentzler RD, Martins RG, Stevenson JP, Jalal SI, Panwalkar A, Yang JC, Gubens M, et al, and KEYNOTE-021 Investigators. Carboplatin and pemetrexed with or without pembrolizumab for advanced, non-squamous non-small-cell lung cancer: a randomised, phase 2 cohort of the open-label KEYNOTE-021 study. *Lancet Oncol.* 2016; 17:1497–508.  
[https://doi.org/10.1016/S1470-2045\(16\)30498-3](https://doi.org/10.1016/S1470-2045(16)30498-3)  
PMID:27745820
5. Reck M, Rodríguez-Abreu D, Robinson AG, Hui R, Csőszi T, Fülöp A, Gottfried M, Peled N, Tafreshi A, Cuffe S, O'Brien M, Rao S, Hotta K, et al, and KEYNOTE-024 Investigators. Pembrolizumab versus Chemotherapy for PD-L1-Positive Non-Small-Cell Lung Cancer. *N Engl J Med.* 2016; 375:1823–33.  
<https://doi.org/10.1056/NEJMoa1606774>  
PMID:27718847
6. Antonia SJ, López-Martin JA, Bendell J, Ott PA, Taylor M, Eder JP, Jäger D, Pietanza MC, Le DT, de Braud F, Morse MA, Ascierto PA, Horn L, et al. Nivolumab alone and nivolumab plus ipilimumab in recurrent small-cell lung cancer (CheckMate 032): a multicentre, open-label, phase 1/2 trial. *Lancet Oncol.* 2016; 17:883–95.  
[https://doi.org/10.1016/S1470-2045\(16\)30098-5](https://doi.org/10.1016/S1470-2045(16)30098-5)  
PMID:27269741
7. Fehrenbacher L, Spira A, Ballinger M, Kowanetz M, Vansteenkiste J, Mazieres J, Park K, Smith D, Artal-Cortes A, Lewanski C, Braiteh F, Waterkamp D, He P, et al, and POPLAR Study Group. Atezolizumab versus docetaxel for patients with previously treated non-small-cell lung cancer (POPLAR): a multicentre, open-label, phase 2 randomised controlled trial. *Lancet.* 2016; 387:1837–46.  
[https://doi.org/10.1016/S0140-6736\(16\)00587-0](https://doi.org/10.1016/S0140-6736(16)00587-0)  
PMID:26970723
8. Herbst RS, Baas P, Kim DW, Felip E, Pérez-Gracia JL, Han JY, Molina J, Kim JH, Arvis CD, Ahn MJ, Majem M, Fidler MJ, de Castro G Jr, et al. Pembrolizumab versus docetaxel for previously treated, PD-L1-



- positive, advanced non-small-cell lung cancer (KEYNOTE-010): a randomised controlled trial. *Lancet*. 2016; 387:1540–50.  
[https://doi.org/10.1016/S0140-6736\(15\)01281-7](https://doi.org/10.1016/S0140-6736(15)01281-7)  
PMID:26712084
9. Borghaei H, Paz-Ares L, Horn L, Spigel DR, Steins M, Ready NE, Chow LQ, Vokes EE, Felip E, Holgado E, Barlesi F, Kohlhäufel M, Arrieta O, et al. Nivolumab versus Docetaxel in Advanced Nonsquamous Non-Small-Cell Lung Cancer. *N Engl J Med*. 2015; 373:1627–39.  
<https://doi.org/10.1056/NEJMoa1507643>  
PMID:26412456
  10. Brahmer J, Reckamp KL, Baas P, Crinò L, Eberhardt WE, Poddubskaya E, Antonia S, Pluzanski A, Vokes EE, Holgado E, Waterhouse D, Ready N, Gainor J, et al. Nivolumab versus Docetaxel in Advanced Squamous-Cell Non-Small-Cell Lung Cancer. *N Engl J Med*. 2015; 373:123–35.  
<https://doi.org/10.1056/NEJMoa1504627>  
PMID:26028407
  11. Gandhi L, Rodríguez-Abreu D, Gadgeel S, Esteban E, Felip E, De Angelis F, Domine M, Clingan P, Hochmair MJ, Powell SF, Cheng SY, Bischoff HG, Peled N, et al, and KEYNOTE-189 Investigators. Pembrolizumab plus Chemotherapy in Metastatic Non-Small-Cell Lung Cancer. *N Engl J Med*. 2018; 378:2078–92.  
<https://doi.org/10.1056/NEJMoa1801005>  
PMID:29658856
  12. Antonia SJ, Villegas A, Daniel D, Vicente D, Murakami S, Hui R, Yokoi T, Chiappori A, Lee KH, de Wit M, Cho BC, Bourhaba M, Quantin X, et al, and PACIFIC Investigators. Durvalumab after Chemoradiotherapy in Stage III Non-Small-Cell Lung Cancer. *N Engl J Med*. 2017; 377:1919–29.  
<https://doi.org/10.1056/NEJMoa1709937>  
PMID:28885881
  13. Paz-Ares L, Dvorkin M, Chen Y, Reinmuth N, Hotta K, Trukhin D, Statsenko G, Hochmair MJ, Özgüroğlu M, Ji JH, Voitko O, Poltoratskiy A, Ponce S, et al, and CASPIAN Investigators. Durvalumab plus platinum-etoposide versus platinum-etoposide in first-line treatment of extensive-stage small-cell lung cancer (CASPIAN): a randomised, controlled, open-label, phase 3 trial. *Lancet*. 2019; 394:1929–39.  
[https://doi.org/10.1016/S0140-6736\(19\)32222-6](https://doi.org/10.1016/S0140-6736(19)32222-6)  
PMID:31590988
  14. Antonia SJ, Villegas A, Daniel D, Vicente D, Murakami S, Hui R, Kurata T, Chiappori A, Lee KH, de Wit M, Cho BC, Bourhaba M, Quantin X, et al, and PACIFIC Investigators. Overall Survival with Durvalumab after Chemoradiotherapy in Stage III NSCLC. *N Engl J Med*. 2018; 379:2342–50.  
<https://doi.org/10.1056/NEJMoa1809697>  
PMID:30280658
  15. Hui R, Özgüroğlu M, Villegas A, Daniel D, Vicente D, Murakami S, Yokoi T, Chiappori A, Lee KH, de Wit M, Cho BC, Gray JE, Rydén A, et al. Patient-reported outcomes with durvalumab after chemoradiotherapy in stage III, unresectable non-small-cell lung cancer (PACIFIC): a randomised, controlled, phase 3 study. *Lancet Oncol*. 2019; 20:1670–80.  
[https://doi.org/10.1016/S1470-2045\(19\)30519-4](https://doi.org/10.1016/S1470-2045(19)30519-4)  
PMID:31601496
  16. Giraldo NA, Becht E, Vano Y, Petitprez F, Lacroix L, Validire P, Sanchez-Salas R, Ingels A, Oudard S, Moatti A, Buttard B, Bourass S, Germain C, et al. Tumor-Infiltrating and Peripheral Blood T-cell Immunophenotypes Predict Early Relapse in Localized Clear Cell Renal Cell Carcinoma. *Clin Cancer Res*. 2017; 23:4416–28.  
<https://doi.org/10.1158/1078-0432.CCR-16-2848>  
PMID:28213366
  17. Nunes C, Wong R, Mason M, Fegan C, Man S, Pepper C. Expansion of a CD8(+)PD-1(+) replicative senescence phenotype in early stage CLL patients is associated with inverted CD4:CD8 ratios and disease progression. *Clin Cancer Res*. 2012; 18:678–87.  
<https://doi.org/10.1158/1078-0432.CCR-11-2630>  
PMID:22190592
  18. Braun DA, Hou Y, Bakouny Z, Ficial M, Sant' Angelo M, Forman J, Ross-Macdonald P, Berger AC, Jegede OA, Elagina L, Steinharter J, Sun M, Wind-Rotolo M, et al. Interplay of somatic alterations and immune infiltration modulates response to PD-1 blockade in advanced clear cell renal cell carcinoma. *Nat Med*. 2020; 26:909–18.  
<https://doi.org/10.1038/s41591-020-0839-y>  
PMID:32472114
  19. Braun DA, Hou Y, Bakouny Z, Ficial M, Sant' Angelo M, Ross-Macdonald P, Jegede O, Sun M, Wind-Rotolo M, Pignon JC, Neuberg DS, Catalano PJ, Freeman GJ, et al. Immunogenomic characterization of advanced clear cell renal cell carcinoma treated with PD-1 blockade. *Journal of Clinical Oncology*. 2020; 38:5010.  
[https://doi.org/10.1200/JCO.2020.38.15\\_suppl.5010](https://doi.org/10.1200/JCO.2020.38.15_suppl.5010)
  20. Ricciuti B, Arbour KC, Lin JJ. Effect of STK11 mutations on efficacy of PD-1 inhibition in non-small cell lung cancer (NSCLC) and dependence on KRAS mutation status. *Journal of Clinical Oncology*. 2020 (Suppl 15); 38:e15113–e15113.  
[https://doi.org/10.1200/JCO.2020.38.15\\_suppl.e15113](https://doi.org/10.1200/JCO.2020.38.15_suppl.e15113)
  21. ICGC/TCGA Pan-Cancer Analysis of Whole Genomes Consortium. Pan-cancer analysis of whole genomes. *Nature*. 2020; 578:82–93.

- <https://doi.org/10.1038/s41586-020-1969-6>  
PMID:[32025007](https://pubmed.ncbi.nlm.nih.gov/32025007/)
22. Pan-Cancer Analysis Pinpoints Targets in PI3K Pathway. *Cancer Discov.* 2017; 7:OF6.  
<https://doi.org/10.1158/2159-8290.CD-NB2017-092>  
PMID:[28619980](https://pubmed.ncbi.nlm.nih.gov/28619980/)
23. Zeng D, Li M, Zhou R, Zhang J, Sun H, Shi M, Bin J, Liao Y, Rao J, Liao W. Tumor Microenvironment Characterization in Gastric Cancer Identifies Prognostic and Immunotherapeutically Relevant Gene Signatures. *Cancer Immunol Res.* 2019; 7:737–50.  
<https://doi.org/10.1158/2326-6066.CIR-18-0436>  
PMID:[30842092](https://pubmed.ncbi.nlm.nih.gov/30842092/)
24. Ma X, Liu Y, Liu Y, Alexandrov LB, Edmonson MN, Gawad C, Zhou X, Li Y, Rusch MC, Easton J, Huether R, Gonzalez-Pena V, Wilkinson MR, et al. Pan-cancer genome and transcriptome analyses of 1,699 paediatric leukaemias and solid tumours. *Nature.* 2018; 555:371–76.  
<https://doi.org/10.1038/nature25795> PMID:[29489755](https://pubmed.ncbi.nlm.nih.gov/29489755/)
25. Tekpli X, Lien T, Røssevold AH, Nebdal D, Borgen E, Ohnstad HO, Kyte JA, Vallon-Christersson J, Fongaard M, Due EU, Svartdal LG, Sveli MA, Garred Ø, et al, and OSBREAC. An independent poor-prognosis subtype of breast cancer defined by a distinct tumor immune microenvironment. *Nat Commun.* 2019; 10:5499.  
<https://doi.org/10.1038/s41467-019-13329-5>  
PMID:[31796750](https://pubmed.ncbi.nlm.nih.gov/31796750/)
26. Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, Hoang CD, Diehn M, Alizadeh AA. Robust enumeration of cell subsets from tissue expression profiles. *Nat Methods.* 2015; 12:453–57.  
<https://doi.org/10.1038/nmeth.3337> PMID:[25822800](https://pubmed.ncbi.nlm.nih.gov/25822800/)
27. Thorsson V, Gibbs DL, Brown SD, Wolf D, Bortone DS, Ou Yang TH, Porta-Pardo E, Gao GF, Plaisier CL, Eddy JA, Ziv E, Culhane AC, Paull EO, et al, and Cancer Genome Atlas Research Network. The Immune Landscape of Cancer. *Immunity.* 2018; 48:812–30.e14.  
<https://doi.org/10.1016/j.immuni.2018.03.023>  
PMID:[29628290](https://pubmed.ncbi.nlm.nih.gov/29628290/)
28. Fleischer T, Tekpli X, Mathelier A, Wang S, Nebdal D, Dhakal HP, Sahlberg KK, Schlichting E, Børresen-Dale AL, Borgen E, Naume B, Eskeland R, Frigessi A, et al, and Oslo Breast Cancer Research Consortium (OSBREAC). DNA methylation at enhancers identifies distinct breast cancer lineages. *Nat Commun.* 2017; 8:1379.  
<https://doi.org/10.1038/s41467-017-00510-x>  
PMID:[29123100](https://pubmed.ncbi.nlm.nih.gov/29123100/)
29. Carlini MJ, Dalurzo MC, Lastiri JM, Smith DE, Vasallo BC, Puricelli LI, Lauría de Cidre LS. Mast cell phenotypes and microvessels in non-small cell lung cancer and its prognostic significance. *Hum Pathol.* 2010; 41:697–705.  
<https://doi.org/10.1016/j.humpath.2009.04.029>  
PMID:[20040391](https://pubmed.ncbi.nlm.nih.gov/20040391/)
30. Kurbatov V, Balayev A, Saffarzadeh A, Heller DR, Boffa DJ, Blasberg JD, Lu J, Khan SA. Digital Inference of Immune Microenvironment Reveals Low-Risk Subtype of Early Lung Adenocarcinoma. *Ann Thorac Surg.* 2020; 109:343–49.  
<https://doi.org/10.1016/j.athoracsur.2019.08.050>  
PMID:[31568747](https://pubmed.ncbi.nlm.nih.gov/31568747/)
31. White ES, Strom SR, Wys NL, Arenberg DA. Non-small cell lung cancer cells induce monocytes to increase expression of angiogenic activity. *J Immunol.* 2001; 166:7549–55.  
<https://doi.org/10.4049/jimmunol.166.12.7549>  
PMID:[11390510](https://pubmed.ncbi.nlm.nih.gov/11390510/)
32. Lin GN, Peng JW, Xiao JJ, Liu DY, Xia ZJ. Prognostic impact of circulating monocytes and lymphocyte-to-monocyte ratio on previously untreated metastatic non-small cell lung cancer patients receiving platinum-based doublet. *Med Oncol.* 2014; 31:70.  
<https://doi.org/10.1007/s12032-014-0070-0>  
PMID:[24927957](https://pubmed.ncbi.nlm.nih.gov/24927957/)
33. Chittezhath M, Dhillon MK, Lim JY, Laoui D, Shalova IN, Teo YL, Chen J, Kamaraj R, Raman L, Lum J, Thamboo TP, Chiong E, Zolezzi F, et al. Molecular profiling reveals a tumor-promoting phenotype of monocytes and macrophages in human cancer progression. *Immunity.* 2014; 41:815–29.  
<https://doi.org/10.1016/j.immuni.2014.09.014>  
PMID:[25453823](https://pubmed.ncbi.nlm.nih.gov/25453823/)
34. Sanford DE, Belt BA, Panni RZ, Mayer A, Deshpande AD, Carpenter D, Mitchem JB, Plambeck-Suess SM, Worley LA, Goetz BD, Wang-Gillam A, Eberlein TJ, Denardo DG, et al. Inflammatory monocyte mobilization decreases patient survival in pancreatic cancer: a role for targeting the CCL2/CCR2 axis. *Clin Cancer Res.* 2013; 19:3404–15.  
<https://doi.org/10.1158/1078-0432.CCR-13-0525>  
PMID:[23653148](https://pubmed.ncbi.nlm.nih.gov/23653148/)
35. McGowan M, Hoven AS, Lund-Iversen M, Solberg S, Helland Å, Hirsch FR, Brustugun OT. PIK3CA mutations as prognostic factor in squamous cell lung carcinoma. *Lung Cancer.* 2017; 103:52–57.  
<https://doi.org/10.1016/j.lungcan.2016.11.018>  
PMID:[28024696](https://pubmed.ncbi.nlm.nih.gov/28024696/)
36. Santaniello A, Napolitano F, Servetto A, De Placido P, Silvestris N, Bianco C, Formisano L, Bianco R. Tumour Microenvironment and Immune Evasion in EGFR Addicted NSCLC: Hurdles and Possibilities. *Cancers (Basel).* 2019; 11:1419.  
<https://doi.org/10.3390/cancers11101419>  
PMID:[31554160](https://pubmed.ncbi.nlm.nih.gov/31554160/)

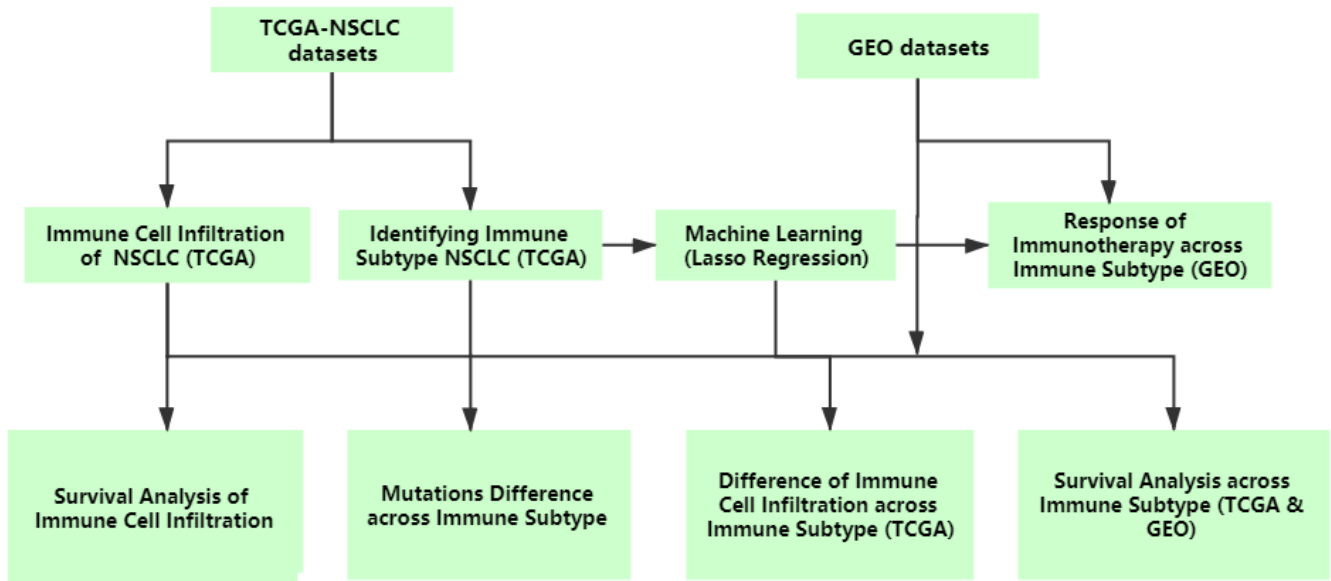
37. Siegel R, Ma J, Zou Z, Jemal A. Cancer statistics, 2014. *CA Cancer J Clin.* 2014; 64:9–29.  
<https://doi.org/10.3322/caac.21208>  
PMID:[24399786](https://pubmed.ncbi.nlm.nih.gov/24399786/)
38. Killock D. Neoadjuvant PD-1 blockade in NSCLC. *Nat Rev Clin Oncol.* 2018; 15:404.  
<https://doi.org/10.1038/s41571-018-0032-0>  
PMID:[29700379](https://pubmed.ncbi.nlm.nih.gov/29700379/)
39. Mok TS, Wu YL, Kudaba I, Kowalski DM, Cho BC, Turna HZ, Castro G Jr, Srimuninnimit V, Laktionov KK, Bondarenko I, Kubota K, Lubiniecki GM, Zhang J, et al, and KEYNOTE-042 Investigators. Pembrolizumab versus chemotherapy for previously untreated, PD-L1-expressing, locally advanced or metastatic non-small-cell lung cancer (KEYNOTE-042): a randomised, open-label, controlled, phase 3 trial. *Lancet.* 2019; 393:1819–30.  
[https://doi.org/10.1016/S0140-6736\(18\)32409-7](https://doi.org/10.1016/S0140-6736(18)32409-7)  
PMID:[30955977](https://pubmed.ncbi.nlm.nih.gov/30955977/)
40. Socinski MA, Jotte RM, Cappuzzo F, Orlandi F, Stroyakovskiy D, Nogami N, Rodríguez-Abreu D, Moro-Sibilot D, Thomas CA, Barlesi F, Finley G, Kelsch C, Lee A, et al, and IMpower150 Study Group. Atezolizumab for First-Line Treatment of Metastatic Nonsquamous NSCLC. *N Engl J Med.* 2018; 378:2288–301.  
<https://doi.org/10.1056/NEJMoa1716948>  
PMID:[29863955](https://pubmed.ncbi.nlm.nih.gov/29863955/)
41. Paz-Ares L, Luft A, Vicente D, Tafreshi A, Gümüş M, Mazières J, Hermes B, Çay Şenler F, Csósz T, Fülöp A, Rodríguez-Cid J, Wilson J, Sugawara S, et al, and KEYNOTE-407 Investigators. Pembrolizumab plus Chemotherapy for Squamous Non-Small-Cell Lung Cancer. *N Engl J Med.* 2018; 379:2040–51.  
<https://doi.org/10.1056/NEJMoa1810865>  
PMID:[30280635](https://pubmed.ncbi.nlm.nih.gov/30280635/)
42. Barlesi F, Vansteenkiste J, Spigel D, Ishii H, Garassino M, de Marinis F, Özgüroğlu M, Szczesna A, Polychronis A, Uslu R, Krzakowski M, Lee JS, Calabrò L, et al. Avelumab versus docetaxel in patients with platinum-treated advanced non-small-cell lung cancer (JAVELIN Lung 200): an open-label, randomised, phase 3 study. *Lancet Oncol.* 2018; 19:1468–79.  
[https://doi.org/10.1016/S1470-2045\(18\)30673-9](https://doi.org/10.1016/S1470-2045(18)30673-9)  
PMID:[30262187](https://pubmed.ncbi.nlm.nih.gov/30262187/)
43. Kim CG, Hong MH, Kim KH, Seo IH, Ahn BC, Pyo KH, Synn CB, Yoon HI, Shim HS, Lee YI, Choi SJ, Lee YJ, Kim EJ, et al. Dynamic changes in circulating PD-1<sup>+</sup>CD8<sup>+</sup> T lymphocytes for predicting treatment response to PD-1 blockade in patients with non-small-cell lung cancer. *Eur J Cancer.* 2021; 143:113–26.  
<https://doi.org/10.1016/j.ejca.2020.10.028>  
PMID:[33302114](https://pubmed.ncbi.nlm.nih.gov/33302114/)
44. Socinski MA, Obasaju C, Gandara D, Hirsch FR, Bonomi P, Bunn P, Kim ES, Langer CJ, Natale RB, Novello S, Paz-Ares L, Pérol M, Reck M, et al. Clinicopathologic Features of Advanced Squamous NSCLC. *J Thorac Oncol.* 2016; 11:1411–22.  
<https://doi.org/10.1016/j.jtho.2016.05.024>  
PMID:[27296106](https://pubmed.ncbi.nlm.nih.gov/27296106/)
45. Lin G, Li C, Li PS, Fang WZ, Xu HP, Gong YH, Zhu ZF, Hu Y, Liang WH, Chu Q, Zhong WZ, Wu L, Wang HJ, et al. Genomic origin and EGFR-TKI treatments of pulmonary adenosquamous carcinoma. *Ann Oncol.* 2020; 31:517–24.  
<https://doi.org/10.1016/j.annonc.2020.01.014>  
PMID:[32151507](https://pubmed.ncbi.nlm.nih.gov/32151507/)
46. Gourd E. Durvalumab boosts progression-free survival in NSCLC. *Lancet Oncol.* 2018; 19:e11.  
[https://doi.org/10.1016/S1470-2045\(17\)30895-1](https://doi.org/10.1016/S1470-2045(17)30895-1)  
PMID:[29175147](https://pubmed.ncbi.nlm.nih.gov/29175147/)
47. Leslie M. Pembrolizumab Outperforms Docetaxel for NSCLC. *Cancer Discov.* 2016; 6:223.  
<https://doi.org/10.1158/2159-8290.CD-NB2016-008>  
PMID:[26811324](https://pubmed.ncbi.nlm.nih.gov/26811324/)
48. Zhu H, Yu Y, Zheng Y, Xu B, Zheng S, Zeng F, Xie W, Huang L, Li F, Lin W, Liao G, Wu S, Liu X, et al. KEAP1/NFE2L2 as a prognostic biomarker on immunotherapy and correlation with immune infiltrates in non-small cell lung cancer (NSCLC). *J Clin Oncol.* 2020; 25:e15.  
[https://doi.org/10.1200/JCO.2020.38.15\\_suppl.e21551](https://doi.org/10.1200/JCO.2020.38.15_suppl.e21551)
49. Zeng J, Wang G, Yan Z, Zheng H, Li J, Tao F. PTPRD/PTPRT mutations as potential positive predictor for response of immune checkpoint inhibitors in non-small cell lung cancer (NSCLC). *J Clin Oncol.* 2020; 25:e15112.  
[https://doi.org/10.1200/JCO.2020.38.15\\_suppl.e15112](https://doi.org/10.1200/JCO.2020.38.15_suppl.e15112)
50. Rochigneux P, Chretien AS, Rossille D, Chanez B, Billon E, Fattori S, Corre R, Pakradouni J, Goncalves A, Fest T, Madroszyk A, Olive D. Soluble BTN2A1 as a potential predictive biomarker of immune checkpoint inhibitor efficacy in advanced non-small cell lung cancer (NSCLC). *J Clin Oncol.* 2020; 25:9561.  
[https://doi.org/10.1200/JCO.2020.38.15\\_suppl.9561](https://doi.org/10.1200/JCO.2020.38.15_suppl.9561)
51. Janning M, Kobus F, Babayan A, Wikman H, Velthaus JL, Bergmann S, Schatz S, Falk M, Berger LA, Böttcher LM, Päsler S, Gorges TM, O’Flaherty L, et al. Determination of PD-L1 Expression in Circulating Tumor Cells of NSCLC Patients and Correlation with Response to PD-1/PD-L1 Inhibitors. *Cancers (Basel).* 2019; 11:835.  
<https://doi.org/10.3390/cancers11060835>  
PMID:[31212989](https://pubmed.ncbi.nlm.nih.gov/31212989/)
52. Abdel-Rahman O. Correlation between PD-L1 expression and outcome of NSCLC patients treated

- with anti-PD-1/PD-L1 agents: A meta-analysis. *Crit Rev Oncol Hematol*. 2016; 101:75–85.  
<https://doi.org/10.1016/j.critrevonc.2016.03.007>  
PMID:[26969107](https://pubmed.ncbi.nlm.nih.gov/26969107/)
53. Choo YW, Kang M, Kim HY, Han J, Kang S, Lee JR, Jeong GJ, Kwon SP, Song SY, Go S, Jung M, Hong J, Kim BS. M1 Macrophage-Derived Nanovesicles Potentiate the Anticancer Efficacy of Immune Checkpoint Inhibitors. *ACS Nano*. 2018; 12:8977–93.  
<https://doi.org/10.1021/acsnano.8b02446>  
PMID:[30133260](https://pubmed.ncbi.nlm.nih.gov/30133260/)
54. Finotello F, Trajanoski Z. New strategies for cancer immunotherapy: targeting regulatory T cells. *Genome Med*. 2017; 9:10.  
<https://doi.org/10.1186/s13073-017-0402-8>  
PMID:[28129791](https://pubmed.ncbi.nlm.nih.gov/28129791/)
55. Borst J, Ahrends T, Bąbała N, Melief CJ, Kastenmüller W. CD4<sup>+</sup> T cell help in cancer immunology and immunotherapy. *Nat Rev Immunol*. 2018; 18:635–47.  
<https://doi.org/10.1038/s41577-018-0044-0>  
PMID:[30057419](https://pubmed.ncbi.nlm.nih.gov/30057419/)
56. Brabletz T, Kalluri R, Nieto MA, Weinberg RA. EMT in cancer. *Nat Rev Cancer*. 2018; 18:128–34.  
<https://doi.org/10.1038/nrc.2017.118>  
PMID:[29326430](https://pubmed.ncbi.nlm.nih.gov/29326430/)
57. Aiello NM, Kang Y. Context-dependent EMT programs in cancer metastasis. *J Exp Med*. 2019; 216:1016–26.  
<https://doi.org/10.1084/jem.20181827>  
PMID:[30975895](https://pubmed.ncbi.nlm.nih.gov/30975895/)
58. Giacomelli AO, Yang X, Lintner RE, McFarland JM, Duby M, Kim J, Howard TP, Takeda DY, Ly SH, Kim E, Gannon HS, Hurlula B, Sharpe T, et al. Mutational processes shape the landscape of TP53 mutations in human cancer. *Nat Genet*. 2018; 50:1381–87.  
<https://doi.org/10.1038/s41588-018-0204-y>  
PMID:[30224644](https://pubmed.ncbi.nlm.nih.gov/30224644/)
59. Aisner DL, Sholl LM, Berry LD, Rossi MR, Chen H, Fujimoto J, Moreira AL, Ramalingam SS, Villaruz LC, Otterson GA, Haura E, Politi K, Glisson B, et al, and LCMC2 Investigators. The Impact of Smoking and TP53 Mutations in Lung Adenocarcinoma Patients with Targetable Mutations-The Lung Cancer Mutation Consortium (LCMC2). *Clin Cancer Res*. 2018; 24:1038–47.  
<https://doi.org/10.1158/1078-0432.CCR-17-2289>  
PMID:[29217530](https://pubmed.ncbi.nlm.nih.gov/29217530/)
60. Cancer Genome Atlas Research Network. Comprehensive genomic characterization of squamous cell lung cancers. *Nature*. 2012; 489:519–25.  
<https://doi.org/10.1038/nature11404> PMID:[22960745](https://pubmed.ncbi.nlm.nih.gov/22960745/)
61. Beroukhi R, Mermel CH, Porter D, Wei G, Raychaudhuri S, Donovan J, Barretina J, Boehm JS, Dobson J, Urashima M, Mc Henry KT, Pinchback RM, Ligon AH, et al. The landscape of somatic copy-number alteration across human cancers. *Nature*. 2010; 463:899–905.  
<https://doi.org/10.1038/nature08822>  
PMID:[20164920](https://pubmed.ncbi.nlm.nih.gov/20164920/)

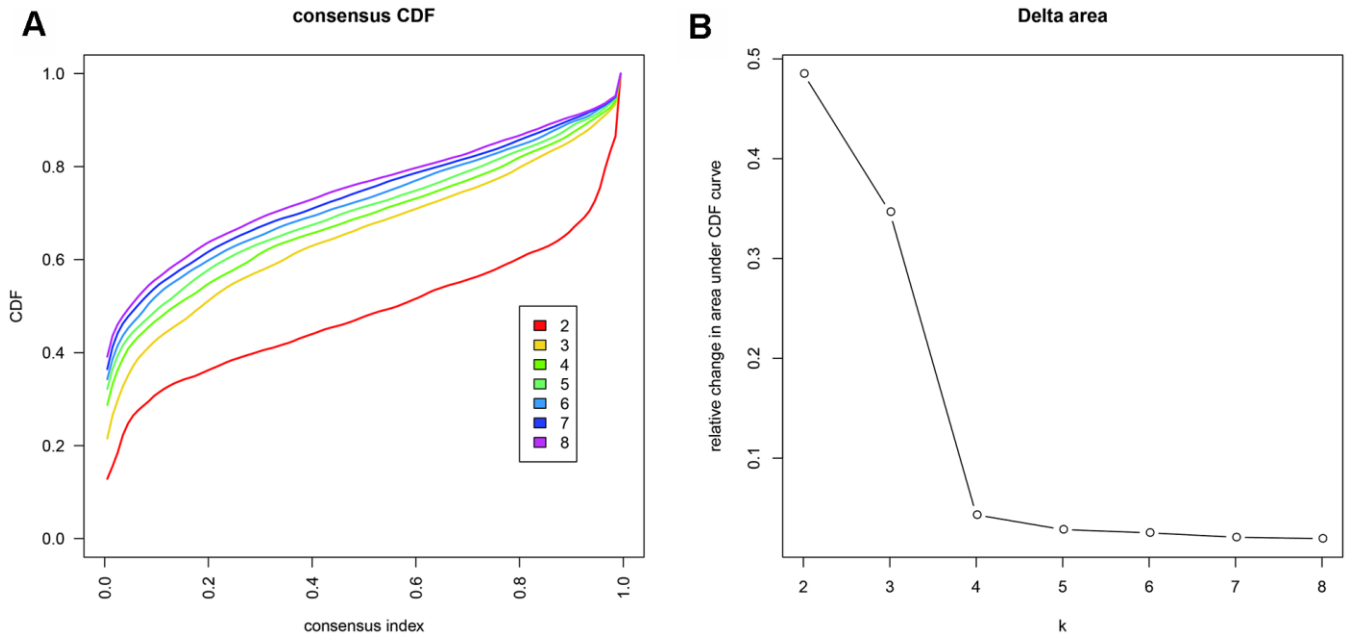


SUPPLEMENTARY MATERIALS

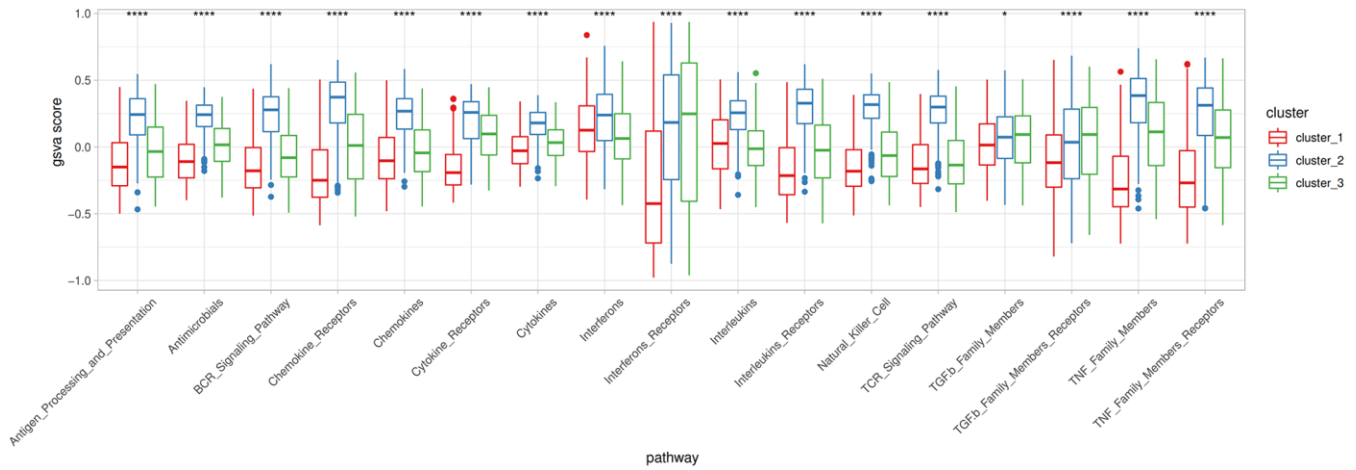
Supplementary Figures



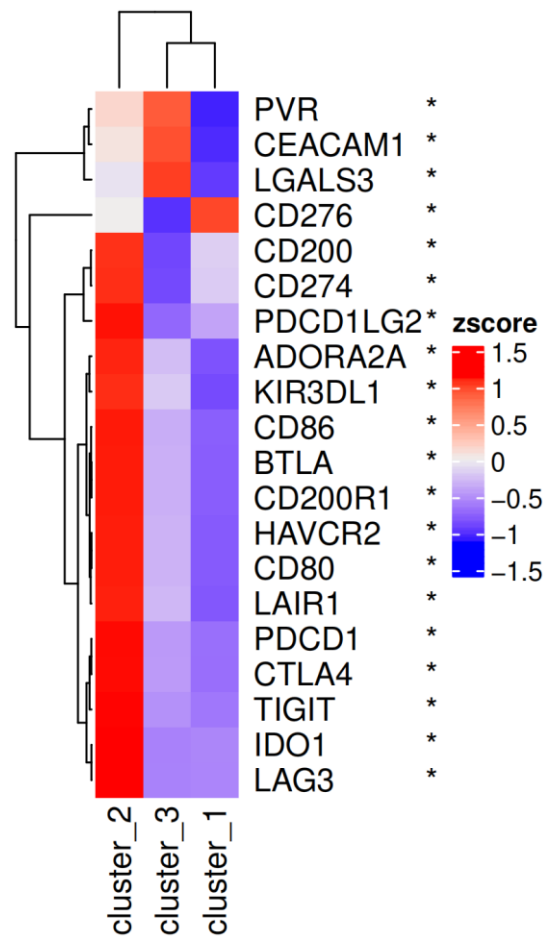
Supplementary Figure 1. Flow diagram of the study.



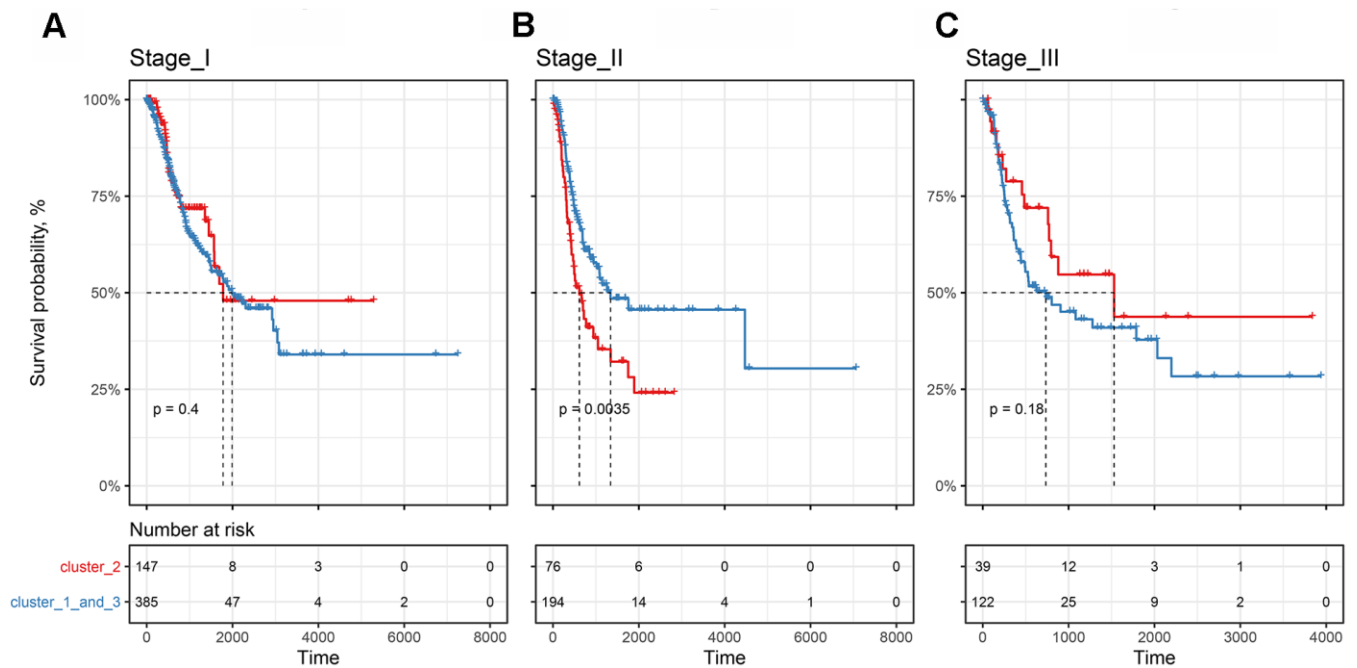
Supplementary Figure 2. Assessment of the optimal number of clusters. (A) The optimization process of Consensus Cumulative Distribution Function (CDF). The abscissa axis represents the consensus index; the ordinate axis represents the CDF value. (B) The change of K value in Delta area. The abscissa axis represents the change of K value, and the ordinate axis represents relative change in area under CDF curve.



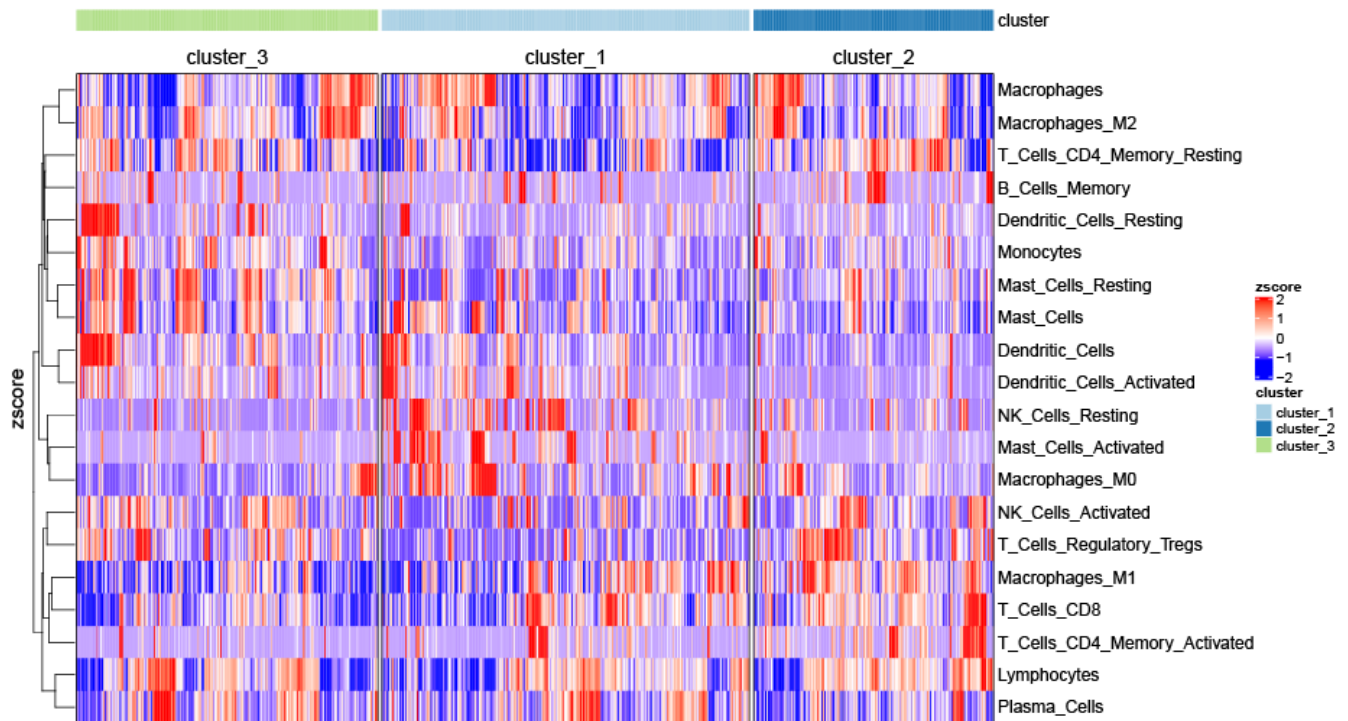
**Supplementary Figure 3. The enrichment scores of immune-related pathways are significantly different in the three clusters: The x-axis is 17 different immune-related pathways, derived from ImmPort.** The y-axis is the pathway enrichment score calculated based on the Gene Set Variation Analysis (GSVA) method. Perform Kruskal test in clusters (\*,  $P < 0.05$ ; \*\*,  $P < 0.01$ ; \*\*\*,  $P < 0.001$ ; \*\*\*\*,  $P < 0.0001$ ).



**Supplementary Figure 4. The average expression of immune checkpoint genes in each cluster: Suppressive immune checkpoints are significantly different in the three clusters.** The expression value is normalized by log2-transformed and zscore. \* Indicates that there is a significant difference in gene expression between the three clusters (Kruskal test, \*,  $P < 0.05$ ).



**Supplementary Figure 5. Survival analysis in different clusters corresponding to different stages: Immunization subgroups can significantly distinguish the progression-free survival of Stage II NSCLC patients from others.** The Stage IV sample size was too small, so it was deleted. The abscissa axis represents survival time, and the ordinate axis represents survival probability. The red curve represents cluster 2, the blue curve represents clusters \_1 and \_3. (A) Survival analysis curve of stage I NSCLC patients. (B) Survival analysis curve of stage II NSCLC patients. (C) Survival analysis curve of stage III NSCLC patients.



**Supplementary Figure 6. Distribution of immune cell infiltration in different clusters.**