Original Article

Development and validation of a gene-based classification model for pN2 lung adenocarcinoma

Jianfei Zhu[1,2#], Wenchen Wang[2#], Yu Ma[3#], Jinbo Zhao[2], Yanlu Xiong[2,4]

[1]Department of Thoracic Surgery, Shaanxi Provincial People's Hospital, Xi'an, China; [2]Department of Thoracic Surgery, Tangdu Hospital, Air Force Medical University, Xi'an, China; [3]Department of Pathology, Shaanxi Provincial People's Hospital, Xi'an, China; [4]Innovation Center for Advanced Medicine, Tangdu Hospital, Air Force Medical University, Xi'an, China

Contributions: (I) Conception and design: J Zhu, Y Xiong; (II) Administrative support: Y Xiong, J Zhao; (III) Provision of study materials or patients: Y Xiong, J Zhao; (IV) Collection and assembly of data: J Zhu, W Wang, Y Ma; (V) Data analysis and interpretation: J Zhu, W Wang, Y Xiong; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

[#]These authors contributed equally to this work.

Correspondence to: Yanlu Xiong. Department of Thoracic Surgery, Tangdu Hospital, Air Force Medical University, Xi'an, China; Innovation Center for Advanced Medicine, Tangdu Hospital, Air Force Medical University, 569 Xinsi Road, Baqiao District, Xi'an 710038, China. Email: xiong21@fmmu.edu.cn; Jinbo Zhao. Department of Thoracic Surgery, Tangdu Hospital, Air Force Medical University, 569 Xinsi Road, Baqiao District, Xi'an 710038, China. Email: zhaojinbo@aliyun.com.

Background: Lung adenocarcinoma (LUAD) with pathological ipsilateral mediastinal lymph node (LN) involvement (pN2) exhibits strong biological and clinical heterogeneity. Thus, it is necessary to classify the biomolecular characteristics that lead to the prognostic heterogeneity of pN2-LUAD.

Methods: The clinical characteristics and bulk RNA sequencing (RNA-seq) data of 75 patients with pN2-LUAD obtained from The Cancer Genome Atlas (TCGA) database were collected as the training set. The disease-free survival (DFS) and overall survival (OS) of patients with different molecular classifications were evaluated. Next, differentially expressed genes (DEGs), biology, and immune cell infiltration in the microenvironment were analysed. Finally, DEGs in the pN2-A and pN2-B groups were included using a least absolute shrinkage and selection operator (LASSO) model, and gene signatures were selected for pN2-A/B type classification. The RNA-seq and single-nucleus RNA sequencing (snRNA-seq) data from our center (n=58) and the GSE68465 dataset (n=53) were used as the validation data sets.

Results: Patients with pN2 LUAD were classified into two distinct molecular categories (pN2-A and pN2-B) based on transcriptome information, pN2-A and pN2-B represent low-risk and high-risk patients, respectively. The survival analysis showed that pN2-A patients had significantly better DFS (P=0.0162) and OS (P=0.0105) compared to pN2-B patients. Multivariate analysis confirmed that molecular classification was an independent factor affecting the prognosis of pN2 LUAD (P=0.0038, and P=0.0024). Next, we found that compared with pN2-A stage patients, pN2-B stage patients had a higher frequency of canonical oncogenic pathway mutations and enrichments. At the single-cell level, we also found that the increase of endothelial cells and the decrease of cytotoxic T/natural killer (NK) cells led to a worse prognosis for pN2-B patients compared to pN2-A patients. Moreover, we established a reasonable gene prediction model of 18 differentially expressed genes (DEGs) to classify the pN2-A and pN2-B patients. Finally, the key above-mentioned results were confirmed using our data and the GES68645 dataset.

Conclusions: The molecular classification of pN2 LUAD is expected to be a powerful supplement to pN2 substaging. Driver gene status and the immune microenvironment mediate different molecular types of LUAD and provide evidence for individualized treatment strategies.

Keywords: Lung adenocarcinoma (LUAD); pN2 stage; prognosis; molecular classification

Submitted Dec 13, 2022. Accepted for publication Mar 21, 2023. Published online Mar 30, 2023.
doi: 10.21037/tlcr-23-16
View this article at: https://dx.doi.org/10.21037/tlcr-23-16

Introduction

The prognosis of non-small-cell lung cancer (NSCLC) with pathological ipsilateral mediastinal lymph node (LN) involvement, defined as stage pN2, is particularly poor. Lung adenocarcinoma (LUAD) has unique biological characteristics and is the most prevalent subtype of NSCLC (1-3). Owing to the use of preoperative and postoperative adjuvant therapies according to driver gene and immune microenvironment status, the prognosis of patients with LUAD has improved considerably (4-6). Unfortunately, the overall outcome is still not satisfactory for all pN2 stage patients, and the 5-year overall survival (OS) rate remains less than 50% (7). Therefore, there is an urgent need to explore the classification of pN2 to predict the prognosis and clarify the mechanism of LUAD LN metastasis, which might help develop targeted interventions to extend the OS of affected patients.

Previous studies have focused on exploring the impact of pathological LN metastasis patterns on the prognosis of patients, and the N staging system for LUAD has been revised periodically to guide clinical practice. In contrast to the LN metastasis of malignant tumours, such as oesophageal cancer and breast cancer, that of LUAD is categorized according to the location of the involved LN and does not consider the number of involved LNs (8,9). In the eighth edition of the revised TNM staging proposed by the International Association for the Study of Lung Cancer (IASLC), the N descriptor is still consistent with the seventh edition due to its high prognostic ability (10). However, the latest staging system (8th edition) recommends considering the addition of the pattern of LN metastasis to pN2 staging, noting the obvious heterogeneity in the pN2 disease. To clarify this prognostic heterogeneity, many studies have proposed various forms of N staging, such as the number of LNs (11), the number of LN stations (12), and the LN metastasis rate (13), while maintaining the characteristic anatomical location classification of LUAD. The previous study indicated that the lncRNAs combined with the above clinical variables was better able to distinguish similar patients (14). However, the genetic variables underlying the aforementioned heterogeneity have not yet been thoroughly explored.

At present, relevant updates on pN staging focus on the LN metastasis patterns, and research on the molecular heterogeneity of pN2 stage disease is limited. The current study aims to evaluate the distinguishing ability and prognostic performance of genetic characteristics identified in The Cancer Genome Atlas (TCGA) database for the pN2 stage of LUAD. Our bulk RNA sequencing (RNA-seq) data and the Gene Expression Omnibus (GEO) database were used to verify our findings. We also mapped the transcriptional landscape of patients with pN2 LUAD using single-nucleus RNA sequencing (snRNA-seq) to screen for vital cell subpopulations that drive its malignant characterization. We present the following article in accordance with the TRIPOD reporting checklist (available at https://tlcr.amegroups.com/article/view/10.21037/tlcr-23-16/rc).

Highlight box

Key findings

- Molecular classification model based on gene signature for pN2 stage LUAD confirmed the underlying reasons for the heterogeneity of this type of disease, and was expected to be a powerful supplement to pN2 substaging.

What is known and what is new?

- In the eighth edition of the pN descriptor of LUAD, for patients with pN2, it was recommended to be subdivided into three categories (pN2a1, pN2a2, and pN2b) according to their prognostic heterogeneity. It indicated that these pN stages still need to be further improved to screen high and low risk patients.
- As far as we know, this study was the first to explore the biological heterogeneity of pN2 patients with LUAD, and divided these patients into two types, pN2-A and pN2-B according to the prognosis.

What is the implication, and what should change now?

- For patients with pN2 stage LUAD, molecular classification model should be added on the basis of traditional TNM staging.

Methods

Transcriptomic and clinical information of public data

The RNA-seq, somatic mutation, and biospecimen clinical data for TCGA-LUAD were downloaded using the TCGA-assembler R package (15,16). After filtering the results, the clinical information of 522 patients, transcriptome data of 515 patients with LUAD, genetic mutation data of 514 patients, and consistent (non-conflicting) prognostic data of 402 patients were available. The transcriptome data and clinical information of GSE68465 (GPL96 platform) were downloaded using the GEO Query R package (17,18); this dataset contained the transcriptome data of 443 patients with primary LUAD and the complete prognostic data of 373 patients.

The clinicopathological characteristics of patients in our center

Patients with LUAD who underwent radical resection from January 2013 to December 2019 in Tangdu Hospital, Air Force Medical University were retrospectively enrolled. The inclusion criteria were as follows: (I) pN2 stage; (II) R0 resection; (III) no other distant organ metastasis; and (IV) specimens suitable for immunohistochemistry (IHC), snRNA-seq, or RNA sequencing. Patients who met the following criteria were excluded from this study: (I) perioperative death; (II) expected survival of fewer than 3 months; (III) underwent neoadjuvant therapy; and (IV) previous history of other malignancies. The last follow-up was conducted on March 31, 2022. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013). The study was approved by the ethics committee of Tangdu Hospital (Approval No. K202003-018), and informed consent was taken from all the patients.

Immunohistochemistry

Briefly, the LUAD samples of 45 pN2 stage patients were dewaxed, antigen repaired, treated with hydrogen peroxide, blocked, incubated with specific primary antibodies (anti-Ki67 antibody: GB111499, Servicebio, Wuhan, China; anti-CLDN5 antibody: GB11290, Servicebio, Wuhan, China; and anti-NKG7 antibody: PA5-67173, Invitrogen, Carlsbad, CA, USA) and horseradish peroxidase (HRP)-labeled secondary antibody, followed by staining, counterstaining, and microscopy.

snRNA-seq

Nuclei were isolated from two frozen LUAD tissues using the Nuclei EZ Lysis buffer containing protease and ribonuclease (RNase) inhibitor (NUC-101, Sigma-Aldrich, St. Louis, MO, USA). Then, the pellet was resuspended and filtered through a 20-mm cell strainer to obtain a suitable cell concentration. Next, refined single-nuclei suspensions were loaded onto 10× Chromium (Princeton, CA, USA) to acquire 8,000 single cells. The subsequent complementary DNA (cDNA) amplification and snRNA-seq libraries were constructed by 10× Genomics Chromium Single Cell 3' kits (V3) as previously described. Further, snRNA-seq libraries were sequenced on an Illumina NovaSeq 6000 (LC Bio Technology Co., Ltd., Hangzhou, China) platform.

Bulk RNA-seq

Total RNA was extracted and prepared from 58 frozen LUAD tissues using TRIzol (Invitrogen, CA, USA) according to the manufacturer's instructions. Total RNA >1.0 µg per sample was used as a starting material for the RNA sample preparation. The captured the messenger RNA (mRNA) was fragmented, and cDNA was synthesized. Subsequently, a sequencing library was generated using uracil-DNA glycosylase (UDG) enzyme (NEB, cat. no. m0280, MA, US) according to the manufacturer's instructions. The library was sequenced on the Illumina NovaSeq 6000 (LC Bio Technology Co., Ltd., Hangzhou, China) platform.

Bioinformatics and statistical methods

For snRNA-seq data, demultiplexing, barcode processing, gene counting, and alignment to the ENSEMBL GRCh38 human reference genome were conducted using Cell Ranger (10× Genomics). The output was then loaded into the Seurat R package for analysis (19). Briefly, after quality control (the number of genes per cell: 200–5,000; the percent of mitochondrial-DNA derived gene: <15%), normalization (Function: NormalizeData), and integration (Function: FindIntegrationAnchors and IntegrateData), the data underwent dimension reduction and clustering (Function: RunPCA, RunUMAP, RunTSNE, and FindClusters). Next, DEGs analysis [Function: FindAllMarkers; P<0.05 and log fold-change (FC) >0.25] and biological process enrichment based on the Kyoto Encyclopedia of Genes and Genomes (KEGG) and the Reactome database (R packages: clusterProfiler and ReactomePA) were performed to identify the cell types. The copy number variation (CNV) was estimated to identify the malignant epithelial cells (R package: infercnv). Cellular communication was evaluated using the CellChat R package.

The bulk transcriptome data were initially standardized. After data normalization using the function [x – min(x)]/[max(x) – min(x)], genes with a higher variation than the upper quartile of the mean absolute deviation (MAD) were regarded as highly variable genes and were scaled by Z transformation for further filtration. The ConsensusClusterPlus package was employed for the clustering analysis (maxK=10, reps=1000, pItem=0.8, clusterAlg= "pam", distance= "Euclidean") (20). Uniform manifold approximation and projection (U-MAP) and T-distribution stochastic neighbour embedding (T-SNE) were used to evaluate the clustering results (21,22). Kaplan-Meier curves were generated to visualize survival probability, the log-rank test was used for statistical analysis [the Benjamini-Hochberg (BH) method was used to correct the p-values for multigroup comparisons], and Cox analysis was applied for risk factor assessment (survival and survminer packages). The DESeq2 package was utilized for DEGs analysis (23,24) under the conditions adj. P<0.05 and absolute value (logFC) >1.

Gene set enrichment analysis (GSEA) was performed to assess biological significance using databases such as Hallmark, KEGG, Reactome, and Gene Ontology (GO) (GSEABase, msigdbr, and clusterProfiler packages) (25,26). The proportion of immune cells was assessed by single-sample GSEA (ssGSEA) (27). Genetic mutation data were processed using the maftools package (28). The wilcox.test function was applied to compare the numerical data, and the chisq.test and fisher.test functions were used for the constituent ratio analysis. Receiver operating characteristic (ROC) curves were used for the diagnostic analyses. The proportion of specific cells in the bulk transcriptome data was assessed by ssGSEA based on the individual gene expression profile obtained from snRNA-seq data. The ggplot2 package was used for graphing. The above operations were performed using the R language (Robert Gentleman, Ross Ihaka, the University of Auckland, New Zealand), with a two-sided P (or adj. P) <0.05 indicating statistical significance (29).

Results

Molecular classification of pN2 stage LUAD

In this study, we obtained the clinical data of 522 patients from TCGA database, including 75 patients with pN2 stage disease, and selected 68 patients for further analysis. Five patients with M1 metastasis, one who was undergoing neoadjuvant treatment, and one with no transcriptome data were excluded. After standardization and normalization, hypervariable genes with a MAD value above the upper quartile were selected (4,959/20,530). After further Z scaling, pN2 stage patients were classified into two molecular types (A and B) by consistent clustering (*Figure 1A*). pN2-A and pN2-B represent low-risk and high-risk patients, respectively. U-MAP and T-SNE dimensionality reduction showed acceptable classification (*Figure 1B,1C*).

We further compared prognostic differences between pN2-A and pN2-B patients. We found that both the disease-free survival (DFS) (30.2 *vs.* 8.2 months, P=0.016) and OS (46.0 *vs.* 15.9 months, P=0.003) of pN2-A patients were significantly better than those of patients with the pN2-B type (*Figure 1D,1E*). These results suggest that pN2 LUAD has marked heterogeneity, necessitating in-depth classification for precise treatment.

Influence of pN2 LUAD molecular classification on clinical decision-making

The different clinical characteristics of the two pN2 types may indicate different clinical strategies. We first compared differences between the pN2-A and pN2-B types in terms of clinicopathological parameters, such as age, sex, pathological stage, and pT category, and only found an association with sex (*Table 1*). We further demonstrated that pN2-A/B type could be an independent risk factor for DFS and OS (*Tables 2,3*).

Patients with disease at different pN stages have distinct differences in prognosis, highlighting the need for diverse clinical strategies as well as the considerable heterogeneity among similar pN stages. We further compared prognostic differences among patients with pN2-A/B, pN0, and pN1 disease to help select the precise clinical treatment. A total of 255 pN0 and 72 pN1 patients with complete prognostic data were selected after excluding those with pM1 disease, neoadjuvant treatment, or no transcriptomic data. We found that pN2-B patients had a markedly shorter DFS than both the pN0 (P<0.001) and pN1 (P=0.019) patients. Also, pN2-A and pN1 patients had a similar DFS (P=0.523). Compared with pN0 patients, pN2-A patients did not exhibit significantly different results (P=0.242), as shown in *Figure 1F*. Similar results were observed for OS (*Figure 1G*).

These results indicated that pN2-A LUAD patients have similar clinical characteristics as pN0/1 LUAD patients. Thus, the clinical treatment of pN2-A LUAD can be conservative, and these patients may benefit more from surgical treatment. However, compared to pN0/1 patients, pN2-B patients require more comprehensive treatment due to their more advanced clinical characteristics.

Biological differences in pN2 LUAD molecular classification

Biological variance in the pN2-A and pN2-B types could possibly explain the clinical differences. We first analyzed the gene mutations in the pN2-A and pN2-B types and found different gene mutation profiles (*Figure 2A,2B*,

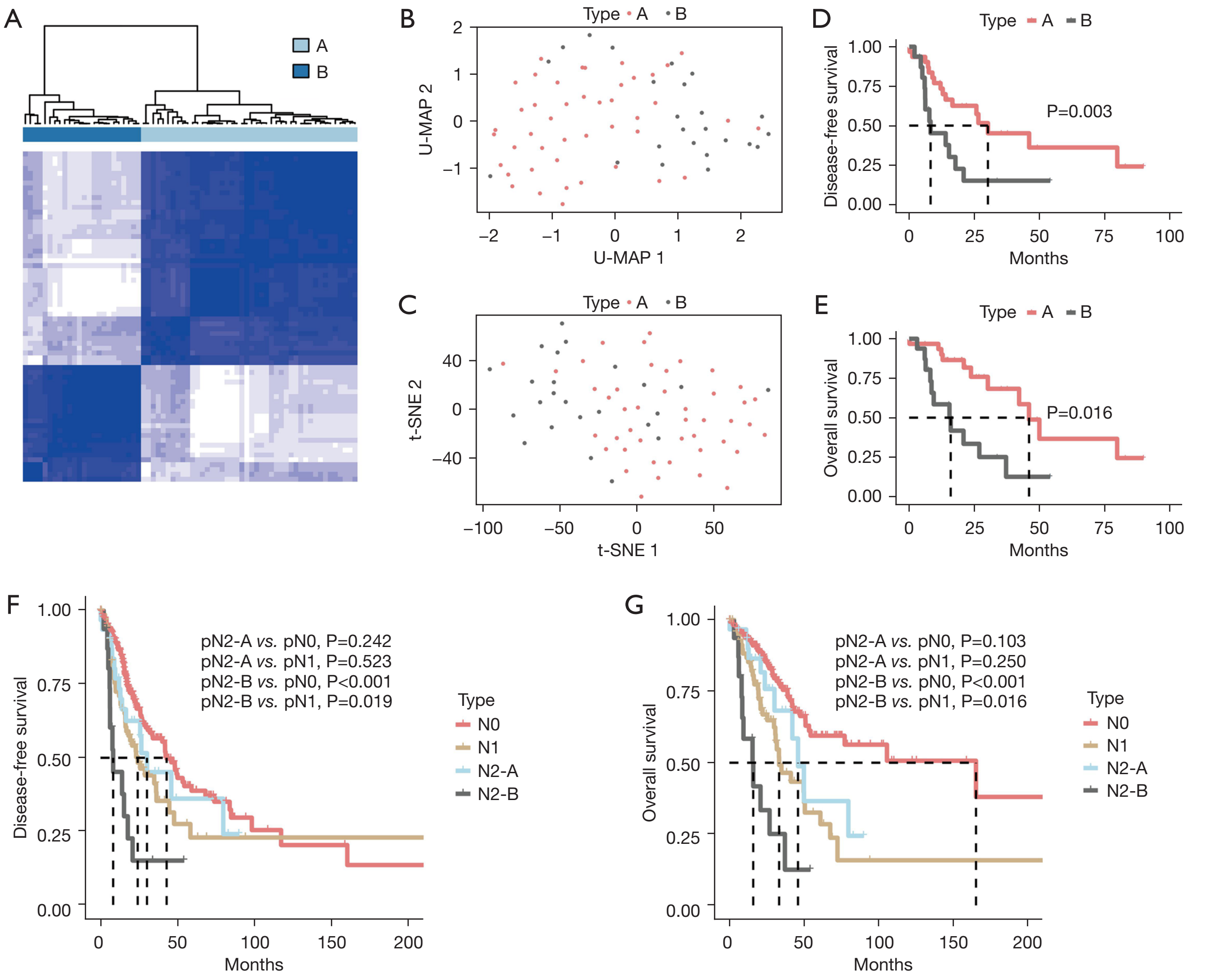

Figure 1 Molecular typing of LUAD with pN2 metastasis in the TCGA-LUAD dataset. (A) Consensus clustering of 68 patients into two types (pN2-A and pN2-B) based on the gene expression profile. (B) U-MAP and (C) t-SNE were applied for dimensionality reduction and visualization of the pN2-A/B molecular types, the red and gray circles represent the gene expression profile of each patient with the two subtypes of pN2-A and pN2-B, respectively. (D) Disease-free survival and (E) overall survival comparison of the pN2-A/B molecular types. (F) Disease-free survival and (G) overall survival comparison and Kaplan-Meier curves for pN0, pN1, pN2-A, and pN2-B LUAD. TCGA, The Cancer Genome Atlas; LUAD, lung adenocarcinoma; U-MAP, uniform manifold approximation and projection; t-SNE, t-distribution stochastic neighbour embedding.

Figure S1A,S1B). Compared with the pN2-A group, the pN2-B group had a higher frequency of canonical oncogenic pathway mutations. Furthermore, we identified 1,843 DEGs between the pN2-B and pN2-A groups (*Figure 2C*). Moreover, we analyzed the gene expression differences in pN2-A or pN2-B tumours compared to N0 or N1 tumours. Compared to N0 or N1 tumours, pN2-B tumours had substantially more DEGs, whereas pN2-A tumours had minor differences in gene expression, possibly suggesting a biological similarity between pN2-A tumours and N0 or N1 tumours. However, a noticeable discrepancy between pN2-B tumours and N0 or N1 tumours was found, which is consistent with the prognostic characteristics mentioned above (*Figure 2D*). Next, we performed GSEA to evaluate the biological behavior of pN2-B and pN2-A tumours. We found that pN2-B LUAD had a higher proliferation rate and classical oncogenic molecular events in three databases (*Figure 2E*: KEGG, *Figure 2F*: Hallmark, *Figure 2G*:

Table 1 Correlation between the pN2 type and clinical variables of TCGA-LUAD cancer cases

Variables	Case (n=68)	pN2 type		P value
		A (n=44)	B (n=24)	
Age (years)				0.4271
>65	37	26	11	
≤65	31	18	13	
Gender				0.0455
Male	30	15	15	
Female	38	29	9	
Smoking history				0.1168
Non-smoker	10	9	1	
Smoker	55	34	21	
NA	3	1	2	
Pathological stage				0.4747
IIIA	63	42	21	
IIIB	5	2	3	
Pathological T category				1
pT1-2	52	34	18	
pT3-4	16	10	6	

TCGA, The Cancer Genome Atlas; LUAD, lung adenocarcinoma; NA, not available.

Table 2 Univariate and multivariate analyses were conducted to screen variables that affect the DFS of patients with pN2 non-small cell lung cancer

Variable	No. (n=47)	DFS (univariate)		DFS (multivariate)	
		HR (95% CI)	P value	HR (95% CI)	P value
Gender		1.4316 (0.6516–3.1453)	0.3717	1.1937 (0.4971–2.8666)	0.6920
Male	16				
Female	31				
Age (years)		0.8745 (0.4142–1.8464)	0.7251	0.8693 (0.3906–1.9347)	0.7315
≤65	22				
>65	25				
Pathological stage		0.4983 (0.1423–1.7449)	0.276	0.7907 (0.1282–4.8763)	0.8003
IIIA	44				
IIIB	3				
Pathological T category		0.6993 (0.2820–1.7342)	0.4402	0.5868 (0.1559–2.2078)	0.4304
pT1-2	39				
pT3-4	8				
N2 type		0.3899 (0.1809–0.8404)	0.0162	0.3356 (0.1455–0.7745)	0.0105
A	31				
B	16				

DFS, disease-free survival; HR, hazard ratio; CI, confidence interval.

Table 3 Univariate and multivariate analyses were performed to screen variables that affect the OS of patients with pN2 non-small cell lung cancer

Variable	No. (n=47)	OS (univariate)		OS (multivariate)	
		HR (95% CI)	P value	HR (95% CI)	P value
Gender		0.8756 (0.3356–2.2845)	0.7860	0.6651 (0.2227–1.9862)	0.4650
Male	16				
Female	31				
Age (years)		1.1907 (0.5102–2.7786)	0.6865	1.2996 (0.5166–3.2695)	0.5777
≤65	22				
>65	25				
Pathological stage		0.4706 (0.1297–1.7073)	0.2517	0.6411 (0.0779–5.2785)	0.6794
IIIA	44				
IIIB	3				
Pathological T category		0.5898 (0.2150–1.6182)	0.3052	0.3531 (0.0662–1.8847)	0.2231
pT1-2	39				
pT3-4	8				
N2 type		0.2779 (0.1168–0.6615)	0.0038	0.2238 (0.0852–0.5882)	0.0024
A	31				
B	16				

OS, overall survival; HR, hazard ratio; CI, confidence interval.

Reactome). Finally, we compared the biological differences of patients with different molecular subtypes of pN2-LUAD by GO analysis (Figure S1C-S1E).

Therefore, the above results suggest that the considerably higher number of oncogenic mutations in pN2-B tumours account for their more aggressive biological behavior, especially unlimited proliferation, thereby determining the severe clinical characteristics. Meanwhile, pN2-A tumours had a similar biological profile to that of pN1 or pN0 tumours, which explains their mild clinical features.

Validation of the pN2 LUAD molecular classification

To confirm the generality of the results, another dataset (GSE68465) was used for verification. As expected, our molecular typing model of pN2 disease is also applicable in the GSE68465 database (53 patients with pN2 disease were classified according to the above procedures) (Figure S2A-S2D). Further GSEA demonstrated that pN2-B tumours exhibited more aggressive malignant behaviors, such as proliferation and classical oncogenic molecular events, than pN2-A tumours (Figure S2E). The clinical analysis demonstrated that pN2-B patients had significantly shorter OS than pN2-A patients, but there was no significant difference in the DFS between the groups (Figure S2F,S2G). Further analysis revealed a shorter OS and DFS in the pN2-B group than in the pN1 (77 patients) or pN0 (245 patients) groups. The pN2-A group had a similar OS and DFS to the pN1 group but poorer OS than the pN0 group (Figure S2H,S2I). We found that pN2-B LUAD had higher proliferation rate and classical oncogenic molecular events in three databases (Figure S3A: Hallmark, Figure S3B: KEGG, and Figure S3C-S3E: GO). Therefore, these results comprehensively verify the robustness of our pN2 LUAD molecular classification.

Characterization of the subpopulation of cancer cells in pN2-LUAD with snRNA-seq data

To comprehensively characterize the cell composition and immune microenvironment of different types of pN2-LUAD, we collected samples from two patients with

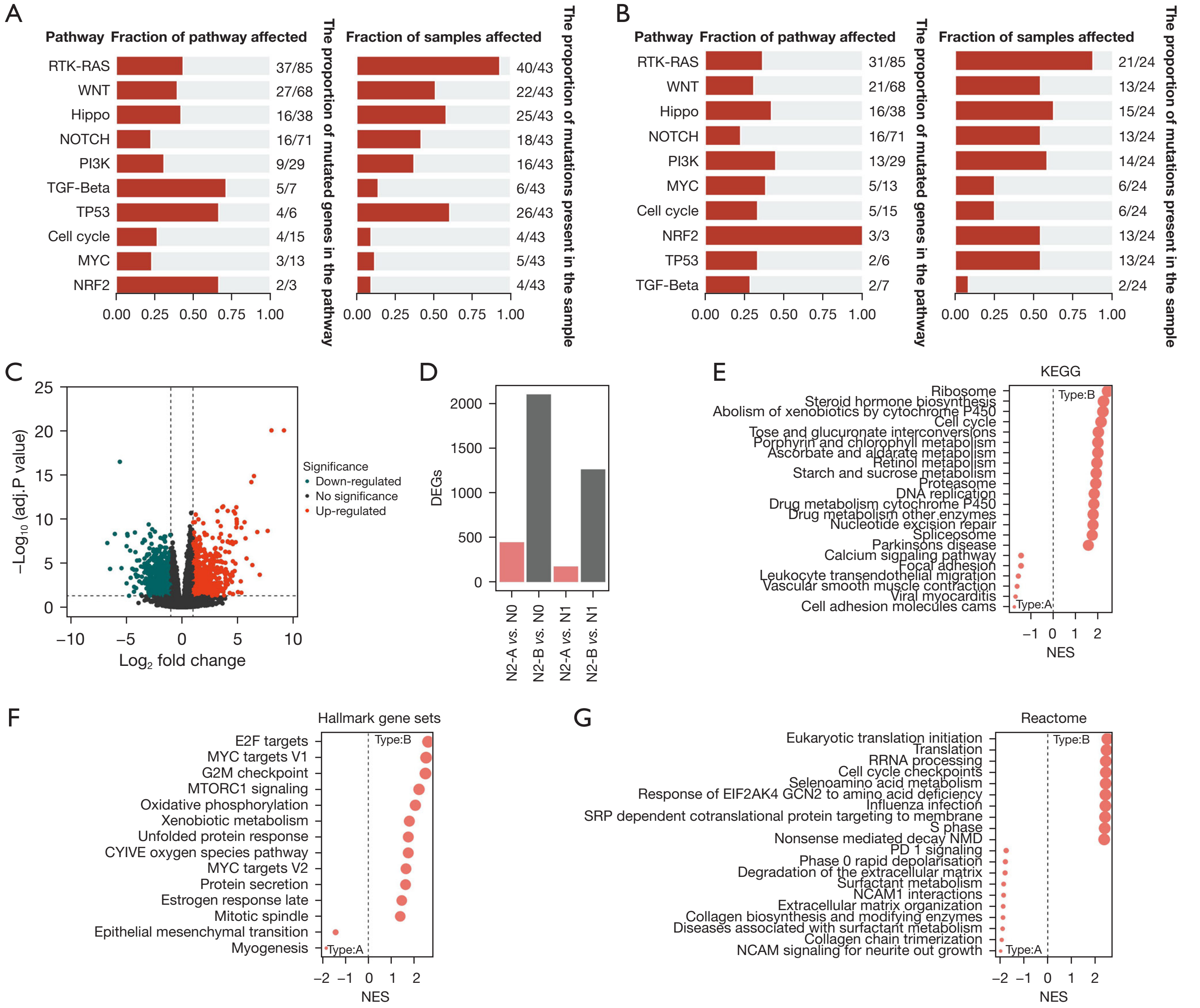

Figure 2 The biological significance of pN2 LUAD molecular typing. (A) pN2-A and (B) pN2-B genetic mutation status in canonical oncogenic pathways, left: the proportion of mutated genes in the pathway, right: the proportion of mutations present in the sample. (C) Volcano plot showing the gene expression differences between the pN2-B and pN2-A groups. (D) Evaluation of DEGs among pN0, pN1, pN2-A, and pN2-B LUAD. (E) KEGG pathway analyses were conducted to assess biological enrichment in the pN2-A/B types. (F) Hallmark and (G) Reactome Pathway Database analysis was performed to assess biological enrichment in the pN2-A/B types by GSEA. LUAD, lung adenocarcinoma; DEGs, differentially expressed genes; KEGG, Kyoto encyclopedia of genes and genomes; GSEA, gene set enrichment analysis; NES, normalized enrichment score.

pN2-A (DFS >60 months) and pN2-B (DFS <12 months), respectively, for snRNA-seq. Finally, a total of 12,019 cells and 41,465 genes passed the quality control. From these, we identified six cell types based on typical cell markers (*Figure 3A,3B*), including epithelial cells, endothelial cells, fibroblasts, T/natural killer (NK) cells, B cells, and myeloid cells. Notably, T/NK cells were enriched in pN2-A patients, while endothelial cells were enriched in pN2-B patients. The possible underlying biological significance of this is explored below.

We first attempted to determine the biological discrepancy of epithelial cells in pN2-A and pN2-B patients.

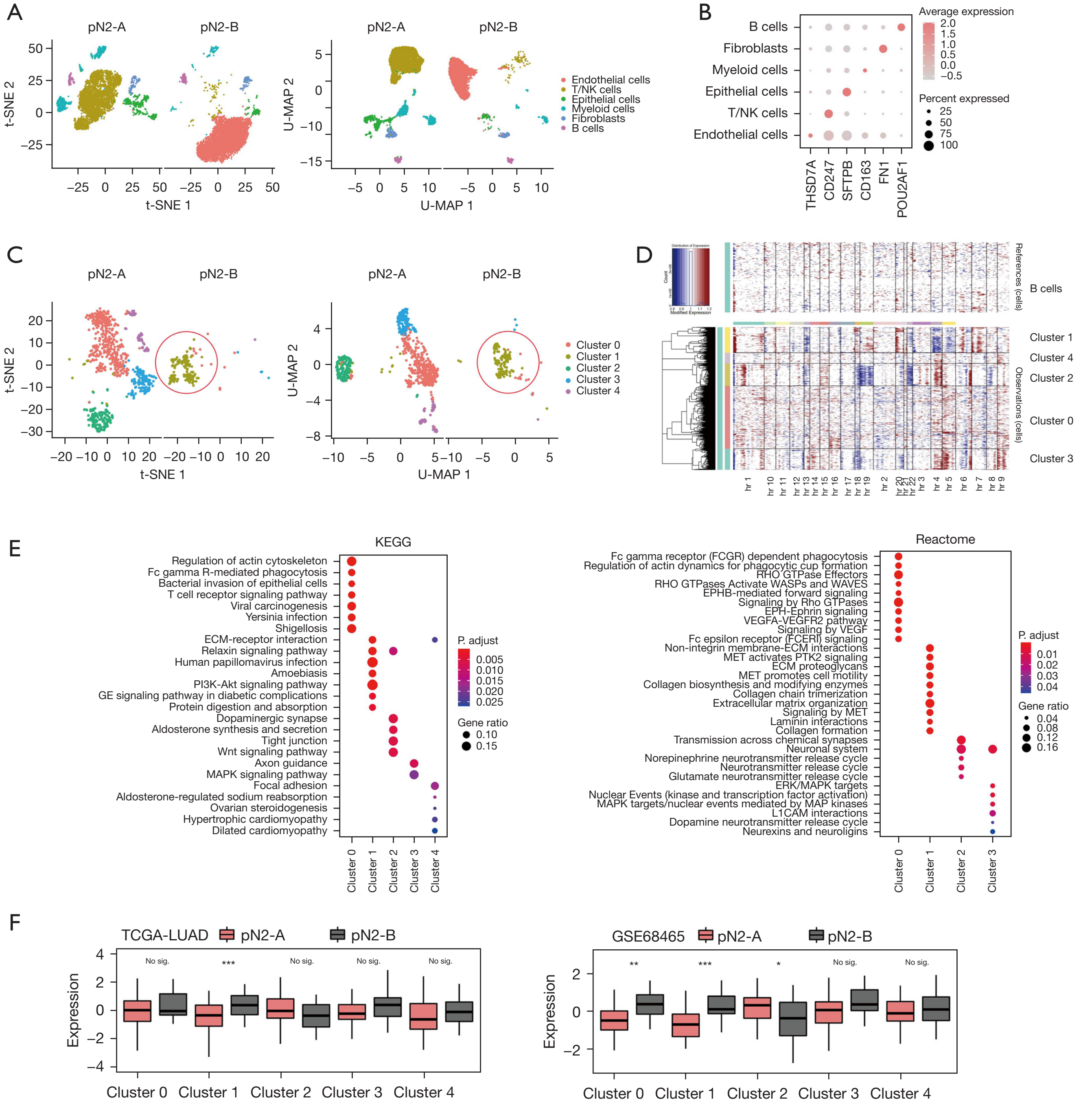

Figure 3 Characteristics of cancer cells and the immune microenvironment in patients with different molecular types of pN2-LUAD at the single-cell level. (A) t-SNE (left panel) and U-MAP (right panel) show the major cell subtypes in patients with pN2-A and pN2-B from the snRNA-seq data. (B) The bubble map shows the marker genes of the six major cell types. (C) t-SNE and U-MAP plots show the clustering of epithelial cells. (D) Normal epithelial cells and cancer cells in different subclusters of epithelial cells were identified by CNVs. (E) KEGG and Reactome analyses were performed to assess the biological enrichment of different subclusters of epithelial cells. (F) The distribution of different epithelial cell subclusters in different molecular types of pN2-LUAD by ssGSEA. LUAD, lung adenocarcinoma; t-SNE, t-distribution stochastic neighbour embedding; U-MAP, Uniform manifold approximation and projection; snRNA-seq, single-nucleus RNA sequencing; CNV, copy number variations; KEGG, Kyoto encyclopedia of genes and genomes; ssGSEA, single-sample gene set enrichment analysis. *, P<0.05; **, P<0.01; ***, P<0.001.

Epithelial cells were clustered into five subtypes (cluster 0–cluster 4); only the cluster 4 subtype was identified as a normal epithelial cell according to its mild CNV, while B cells were used as a reference (*Figure 3C*,*3D*). Interestingly, the cancer cells of cluster 1 were mainly enriched in patients with pN2-B, suggesting that the enrichment of such cancer cells (cluster 1) may be the underlying reason for the aggressive biological behavior of these tumours. Moreover, through pathway enrichment analysis, we found that the cluster 1 cancer cells were exclusively activated in the following signaling pathways: extracellular matrix (ECM), phosphatidylinositol 3 kinase-protein kinase B (PI3K-PKB), and mesenchymal epithelial transition (MET), which are the classical cancer-promoting signaling pathways (*Figure 3E*). We then performed ssGSEA using TCGA-LUAD and GSE68465 data to evaluate the distribution of four cancer cell subpopulations in pN2-A and pN2-B and found that the cancer cells of cluster 1 were specifically enriched in pN2-B patients (*Figure 3F*).

Immunotoxicity of T/NK infiltration can indicate better prognosis in pN2-A patients

To explore the role of different T/NK cell subtypes in pN2-LUAD molecular typing, we clustered T/NK cells into five subtypes, including 2 subtypes of Helper CD4$^+$ T cells (T/NK-C0 and T/NK-C3), 2 subtypes of cytotoxic immune cells (CD8$^+$ T or NK) (T/NK-C1 and T/NK-C4), and 1 subtype of Tregs (T/NK-C2), as shown in *Figure 4A*. Pathway enrichment analysis confirmed that cytotoxic immune cells (T/NK-C1 and T/NK-C4) mainly activated the T-cell receptor and chemokine signaling pathways to exert an immune surveillance function (*Figure 4B*). The cell-cell interaction network also confirmed that cancer cells primarily communicate with cytotoxic immune cells rather than Treg cells (*Figure 4C*). Finally, using IHC, we found that compared with pN2-A, patients with pN2-B had more endothelial cells and fewer NK cells (*Figure 4D*). Based on the above results, we believe that the increase of endothelial cells and the decrease of cytotoxic T/NK cells leads to a worse prognosis for pN2-B patients, as compared to pN2-A patients.

Gene signature for evaluating the pN2 LUAD molecular classification

Finally, we attempted to identify a gene signature to distinguish the pN2-A and pN2-B LUAD types. We selected 68 pN2-A and pN2-B cases from TCGA-LUAD as the training set. We then entered 466 upregulated DEGs and 501 downregulated DEGs between pN2-A and pN2-B into the least absolute shrinkage and selection operator (LASSO) regression model following Z standardization. Based on Lambda.1se, 18 genes were finally selected (*Figure 5A*,*5B*), and the gene signature for pN2-A/B LUAD classification was established:

−0.9173 + *CDK1* × 0.0446 + *CKS1B* × 0.0226 + *GSR* × 0.1064 + *HMMR* × 0.1041 + *ICAM1* × −0.1041 + *NEK2* × 0.3716 + *RRAD* × −0.0559 + *UGCG* × 0.4913 + *VDAC3* × 0.2102 + *B4GALT4* × 0.0526 + *KIF20B* × 0.1609 + *ADAMTSL2* × −0.1019 + *CHL1* × 0.4429 + *DKK1* × 0.1489 + *GABARAPL1* × 0.0110 + *NUSAP1* × 0.2600 + *PLXNA3* × −0.3167 + *SLC38A1* × 0.3819.

The ROC curve for prediction in the training set was highly consistent (*Figure 5C*), it demonstrated that the molecular model had a satisfactory prediction accuracy of LUAD patient prognosis, with an AUC of 1.0. We used 53 pN2 cases in GSE68465 as a validation set and found that the 18-gene signature could effectively distinguish the pN2-A and pN2-B LUAD types (*Figure 5D*). Moreover, our data confirmed that the predictive effect of this model was reasonable (*Figure 5E*).

To verify the accuracy and applicability of the pN2 LUAD classification in the public datasets, our data were also used for verification (58 patients with pN2 stage disease were selected for RNA-seq; two patients with M1 metastasis and one patient who died during the perioperative period were excluded from further analysis). After applying the same procedures, we found the stable, clear, and reliable classification of pN2-A and pN2-B in our dataset (*Figure 6A*). U-MAP and T-SNE also revealed distinct differences in the pN2-A and pN2-B clusters (*Figure 6B*,*6C*). The survival analysis showed that molecular typing could clearly differentiate DFS (P=0.019) and OS (P=0.179) in patients with pN2 disease (*Figure 6D*,*6E*). We also observed similar biological differences between these two groups (Figure S4).

In summary, through the analysis of three databases, we found that the molecular typing pattern based on the gene signature is an important tool for the prognostic prediction of N2 LUAD patients, suggesting that it can be an important complement to traditional TNM staging.

Discussion

The pN2 stage of LUAD was retained from the seventh edition of the TNM staging system to the recently adopted eighth edition. There were no changes in the N

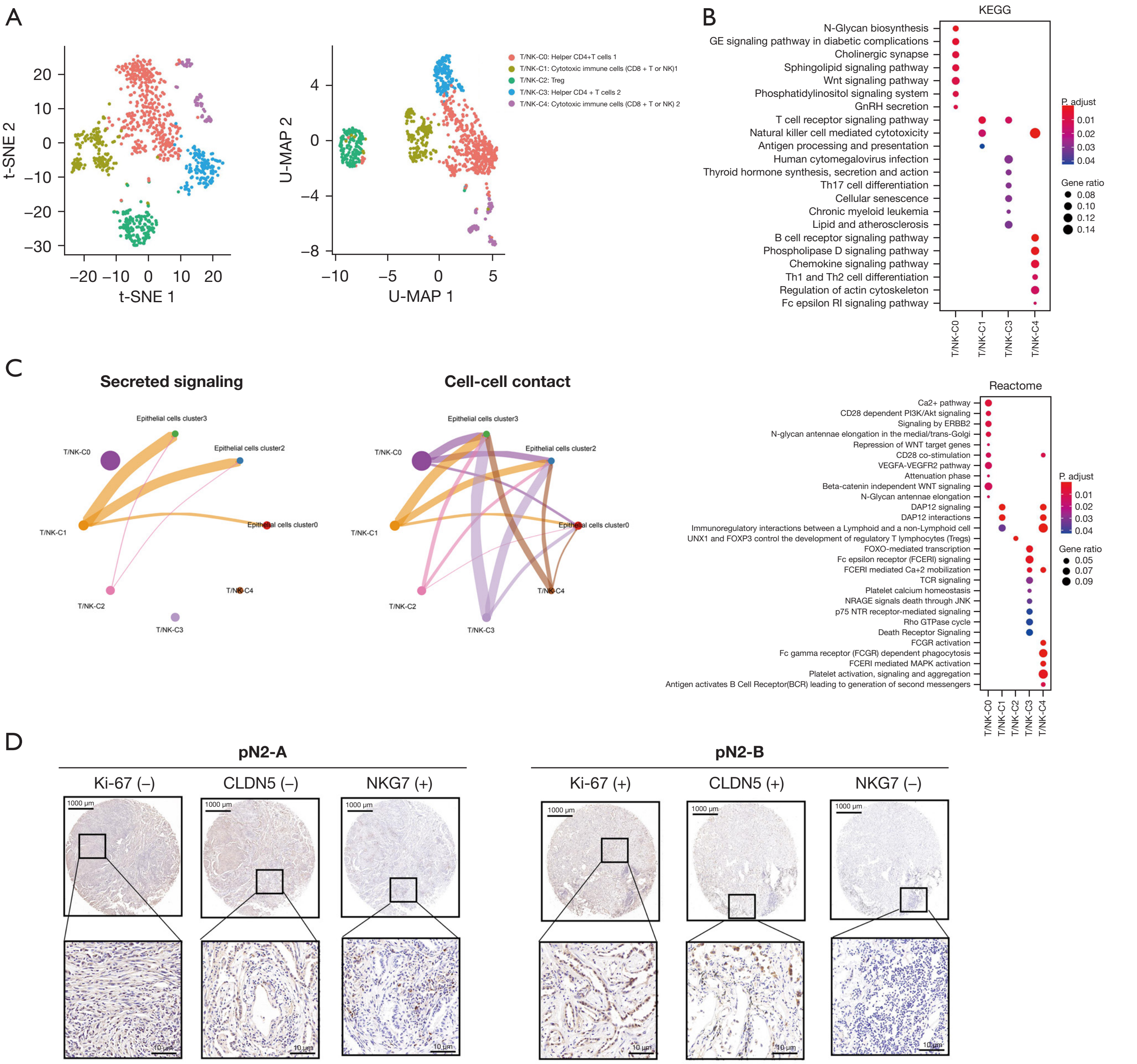

Figure 4 T/NK cell infiltration mediated the biological behavior of pN2-LUAD in the single-cell analysis. (A) t-SNE (left panel) and U-MAP (right panel) show five subclusters of T/NK cells from the snRNA-seq data. (B) KEGG and Reactome analyses were performed to assess the biological enrichment of different subclusters of T/NK cells. (C) Interactions between cancer cells and T/NK cell subclusters. (D) IHC verified the infiltration of immune microenvironment cells in patients with different molecular subtypes of pN2-LUAD. NK, natural killer; LUAD, lung adenocarcinoma; t-SNE, t-distribution stochastic neighbour embedding; U-MAP, uniform manifold approximation and projection; KEGG, Kyoto encyclopedia of genes and genomes; IHC, immunohistochemistry.

descriptors, and anatomical location-based staging is still recommended (1). However, the new staging system proposes that the number of LN stations (one or multiple) and the metastasis mode (the absence or presence of skip metastasis) should be considered in pN2 staging, as it may provide a more accurate prognosis, and that pN2 should be subdivided into pN2a1 (N2+, N1−), pN2a2 (N2+, N1+), and pN2b (multiple N2). Therefore, several variables,

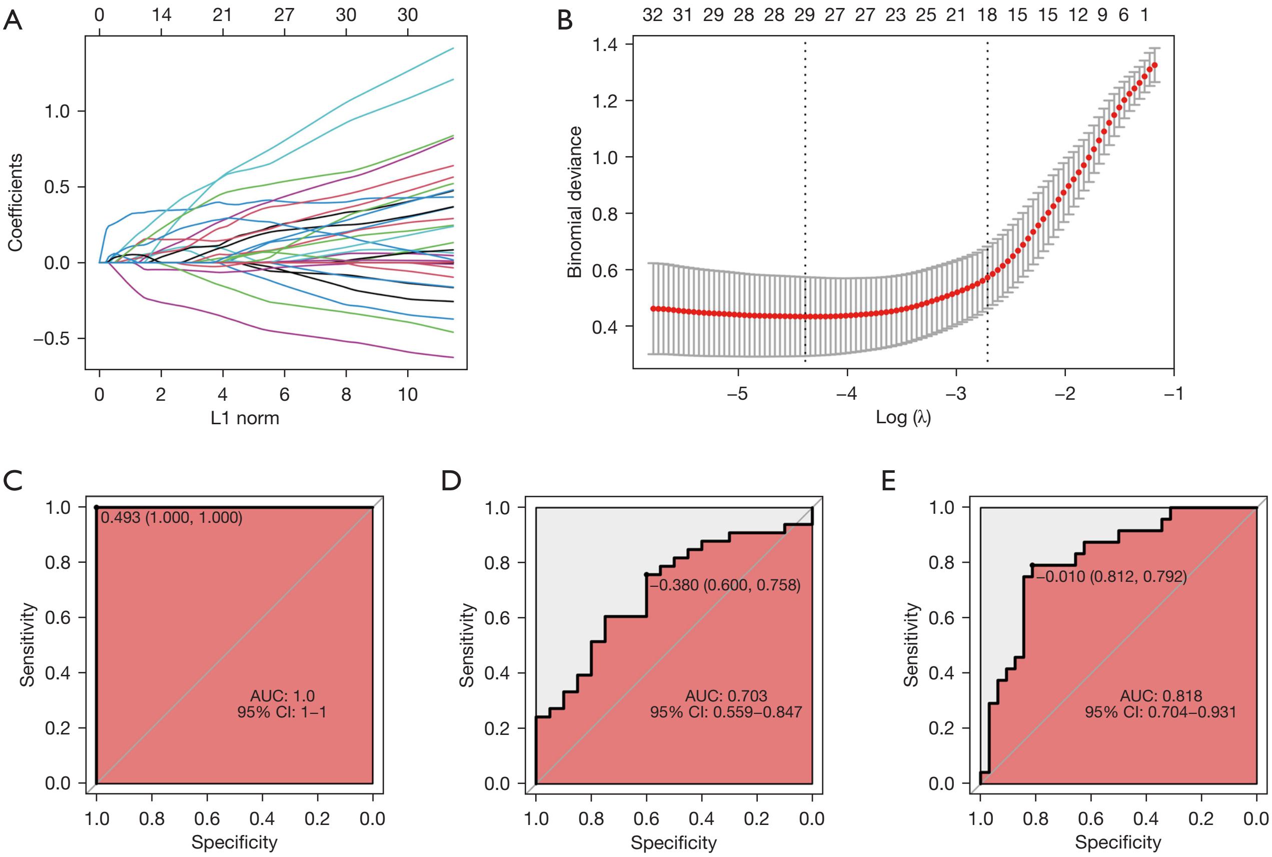

Figure 5 Gene signature for the molecular classification of LUAD with pN2 metastasis. (A) Coefficients and (B) binomial deviations for variables in the LASSO regression model. (C) ROC prediction in the training set of TCGA-LUAD; (D) ROC prediction in the validation set of the GSE68465 dataset; (E) ROC prediction in the validation set of our dataset. LUAD, lung adenocarcinoma; LASSO, least absolute shrinkage and selection operator; ROC, receiver operating characteristic; AUC, area under the curve.

such as the location of LN metastasis, the metastatic LN stations, and the presence of skip metastasis, were incorporated to allow for the detailed stratification of pN2. To the best of our knowledge, this study was the first to explore the biological heterogeneity of patients with pN2 LUAD and to divide these patients into two types, pN2-A and pN2-B, according to prognosis. Notably, in our study, although there was no significant difference in the baseline characteristics, such as age, smoking status, or T stage, between the pN2-A and pN2-B groups but the survival outcomes of LUAD patients were different.

Relevant data suggests that the prognosis of patients with pN2 LUAD is heterogeneous (12,30). Thus, further subcategorization of pN2 disease is crucial for the treatment of LUAD. A study that enrolled 3,791 patients confirmed that zone-based classification and station-based classification had similar effects on the prognostic stratification of pN2 NSCLC. In this previous study, pN2 was divided into four zones: the upper mediastinum zone, the aortopulmonary zone, the subcarinal zone, and the lower mediastinum zone. The researchers found that compared to the station-based model, the zone-based model exhibited a better model fit and comparable prediction errors (31). Moreover, Wang *et al.* (32) researched the role of skip N2 metastasis in operable NSCLC and found that it was a good prognostic factor. Based on the above research results, the eighth edition of TNM staging recommends subdividing pN2 disease; however, the role of this subdivision has not yet been validated. Therefore, Park *et al.* (33) designed a study to verify the recommended change in the N descriptor proposed by IASLC. Cox multivariate analysis showed statistically significant differences in the survival and recurrence rates between pN1a (single station) and pN1b (multiple stations) and between pN2a1/2 and pN2b.

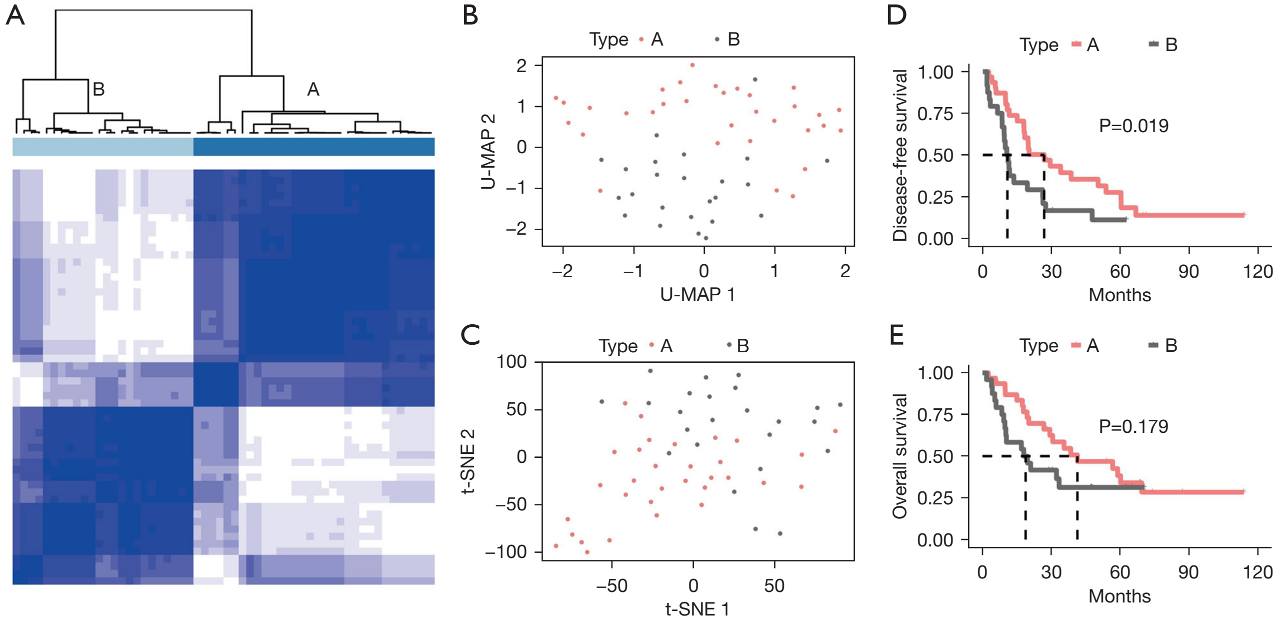

Figure 6 Validation of the molecular typing of LUAD with pN2 metastasis and the clinical and biological assessments from our dataset. (A) Consensus clustering of 55 patients into the pN2-A and pN2-B types based on the gene expression profile. (B) U-MAP and (C) t-SNE were used to evaluate the effects of classification into the pN2-A/B molecular types, the red and gray circles represent the gene expression profile of each patient with the two subtypes of pN2-A and pN2-B, respectively. Comparisons of the (D) disease-free survival and (E) overall survival among the pN2-A/B molecular groups. LUAD, lung adenocarcinoma; U-MAP, uniform manifold approximation and projection; t-SNE, t-distribution stochastic neighbour embedding.

However, N2a1 was not sufficiently different from N1a or N1b. For OS, the difference between N1b and N2a1 was insignificant. Thus, the recommended changes still need to be further improved and updated. Notably, the molecular classification system for the pN2 stage divided this stage into the pN2-A (low-risk) and pN2-B (high-risk) subtypes. Further analysis showed that the prognosis of pN2-A patients was the same as those of N1 or even N0 patients, confirming that molecular typing could be used as a supplement to TNM staging.

Next, we explored the molecular mechanisms underlying the two subtypes of pN2. Compared with pN2-A, pN2-B had a higher frequency of canonical oncogenic pathway mutations, and the results of this study were consistent with those of previous related studies. Liu *et al.* (34) analyzed the correlation of driver gene mutations with clinical T1 stage NSCLC and LN metastasis, and their results confirmed that the incidence of LN metastasis was significantly higher in the mutant group than in the wild-type group (45.1% *vs.* 19.3%, P<0.05). The proportion of vascular invasion and the LN-positive rate were also higher in patients with fusion genes. A retrospective study involving 1,512 cases of LUAD confirmed that there was no difference in recurrence-free survival between the epidermal growth factor receptor (*EGFR*) mutation group and the wild-type group (P=0.266). However, subgroup analysis showed that *EGFR* status played a predictive prognostic role in histological manifestations of acinar pattern adenocarcinoma, papillary adenocarcinoma, adenocarcinoma, and invasive mucinous adenocarcinoma (P=0.015). Further analysis showed that compared with the wild-type cohort, the *EGFR*-mutant cohort had significantly more brain (P=0.004) and bone (P=0.011) metastases (35). By mapping the cells in the tumor microenvironment (TME) and their possible functions, and highlighting the interaction between cancer cells and TME cells, we revealed the cell-cell communication of pN2-LUAD patients with different molecular types at the single-cell level. Notably, T/NK cells were enriched in pN2-A patients, while endothelial cells were enriched in pN2-B patients. GESA confirmed that cytotoxic immune cells (T/NK-C1 and T/NK-C4) mainly activated T-cell receptor and chemokine signaling pathways to exert an immune surveillance function. Re-recruitment of NK cells is expected to be an ideal therapeutic model for pN2-B patients.

Finally, we established a molecular classification model comprising 18 DEGs to distinguish pN2-A and pN2-B LUAD and verified the efficacy of the model in an independent GSE dataset. Many of the selected genes have been confirmed to be involved in the occurrence and development of NSCLC in previous studies but some lack relevant reports. *CDC28* protein kinase regulatory subunit 1 (*CKS1*) is necessary for scf-skp2-mediated p27Kip1 ubiquitination and degradation, which are important steps in the G1/S transition (36,37). Never-in-mitosis-A-related kinase 2 (*NEK2*) is a cell cycle-regulating protein kinase, Chen *et al.* (38) reported that *NEK2* is an oncogene regulated by *EGFR* mutations and is involved in disease progression and therapeutic response in *EGFR* mutated NSCLC. Further study is needed to determine the role of the remaining genes in NSCLC.

This study inevitably has some limitations that should be noted and considered. Firstly, although we chose two databases, TCGA and GEO, the number of patients suitable for further analysis was still limited, and large sample studies must be designed to verify our findings. Secondly, we could not compare our molecular classification with the recommended change in the pN2 descriptor proposed by IASLC due to the data source. Finally, the relevant genes screened in this study still need to be verified in both *in vivo* and *in vitro* experiments.

Conclusions

In addition to heterogeneity in clinicopathological features, the molecular characteristics of pN2 LUAD exhibit obvious heterogeneity. We divided this disease into two distinct molecular types (pN2-A and pN2-B) relevant to prognosis, suggesting that molecular classification can be used as an important supplement to standard LUAD classification. Moreover, driver gene status and the immune microenvironment mediate different molecular types of LUAD and provide evidence for the individualized treatment of LUAD patients during the perioperative period. Finally, our multi-gene molecular feature model could be used as an important tool for pN2 LUAD risk screening.

Acknowledgments

A part of this study was reported in the form of a mini-oral presentation at the 2022 World Conference on Lung Cancer in Vienna, Austria (No. MA06.07; https://wclc2022.iaslc.org/wp-content/uploads/2022/07/WCLC2022-Abstract-Book.pdf), and the abstract as a preprint published on SSRN.

Funding: This study was funded by the National Natural Science Foundation of China (No. 82002421), the Young Talent Program of Tangdu Hospital, the Incubator Fund for Development of Science and Technology of Shaanxi Provincial People's Hospital (No. 2022YJY-26), and the Special Fund for Elite Talents of Shaanxi Provincial People's Hospital (No. 2021JY-20).

Footnote

Reporting Checklist: The authors have completed the TRIPOD reporting checklist. Available at https://tlcr.amegroups.com/article/view/10.21037/tlcr-23-16/rc

Data Sharing Statement: Available at https://tlcr.amegroups.com/article/view/10.21037/tlcr-23-16/dss

Peer Review File: Available at https://tlcr.amegroups.com/article/view/10.21037/tlcr-23-16/prf

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at https://tlcr.amegroups.com/article/view/10.21037/tlcr-23-16/coif). The authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013). The study was approved by the ethics committee of Tangdu Hospital (Approval No. K202003-018), and informed consent was taken from all the patients.

References

1. Asamura H, Chansky K, Crowley J, et al. The International Association for the Study of Lung Cancer Lung Cancer Staging Project: Proposals for the Revision of the N Descriptors in the Forthcoming 8th Edition of the TNM Classification for Lung Cancer. J Thorac Oncol 2015;10:1675-84.
2. Yu Y, Jian H, Shen L, Zhu L, Lu S. Lymph node involvement influenced by lung adenocarcinoma subtypes in tumor size ≤3 cm disease: A study of 2268 cases. Eur J Surg Oncol 2016;42:1714-9.
3. Chen D, Ding Q, Wang W, et al. Characterization of Extracapsular Lymph Node Involvement and Its Clinicopathological Characteristics in Stage II-IIIA Lung Adenocarcinoma. Ann Surg Oncol 2021;28:2088-98.
4. Yue D, Xu S, Wang Q, et al. Erlotinib versus vinorelbine plus cisplatin as adjuvant therapy in Chinese patients with stage IIIA EGFR mutation-positive non-small-cell lung cancer (EVAN): a randomised, open-label, phase 2 trial. Lancet Respir Med 2018;6:863-73.
5. Zhong WZ, Wang Q, Mao WM, et al. Gefitinib Versus Vinorelbine Plus Cisplatin as Adjuvant Treatment for Stage II-IIIA (N1-N2) EGFR-Mutant NSCLC: Final Overall Survival Analysis of CTONG1104 Phase III Trial. J Clin Oncol 2021;39:713-22.
6. Gainor JF. Adjuvant PD-L1 blockade in non-small-cell lung cancer. Lancet 2021;398:1281-3.
7. Deng W, Xu T, Xu Y, et al. Survival Patterns for Patients with Resected N2 Non-Small Cell Lung Cancer and Postoperative Radiotherapy: A Prognostic Scoring Model and Heat Map Approach. J Thorac Oncol 2018;13:1968-74.
8. Gaur P, Sepesi B, Hofstetter WL, et al. A clinical nomogram predicting pathologic lymph node involvement in esophageal cancer patients. Ann Surg 2010;252:611-7.
9. Hessler LK, Molitoris JK, Rosenblatt PY, et al. Factors Influencing Management and Outcome in Patients with Occult Breast Cancer with Axillary Lymph Node Involvement: Analysis of the National Cancer Database. Ann Surg Oncol 2017;24:2907-14.
10. Li S, Yan S, Lu F, et al. Validation of the 8th Edition Nodal Staging and Proposal of New Nodal Categories for Future Editions of the TNM Classification of Non-Small Cell Lung Cancer. Ann Surg Oncol 2021;28:4510-6.
11. Saji H, Tsuboi M, Shimada Y, et al. A proposal for combination of total number and anatomical location of involved lymph nodes for nodal classification in non-small cell lung cancer. Chest 2013;143:1618-25.
12. Legras A, Mordant P, Arame A, et al. Long-term survival of patients with pN2 lung cancer according to the pattern of lymphatic spread. Ann Thorac Surg 2014;97:1156-62.
13. Ding X, Hui Z, Dai H, et al. A Proposal for Combination of Lymph Node Ratio and Anatomic Location of Involved Lymph Nodes for Nodal Classification in Non-Small Cell Lung Cancer. J Thorac Oncol 2016;11:1565-73.
14. Hao X, Li W, Li W, et al. Re-evaluating the need for mediastinal lymph node dissection and exploring lncRNAs as biomarkers of N2 metastasis in T1 lung adenocarcinoma. Transl Lung Cancer Res 2022;11:1079-88.
15. Wei L, Jin Z, Yang S, et al. TCGA-assembler 2: software pipeline for retrieval and processing of TCGA/CPTAC data. Bioinformatics 2018;34:1615-7.
16. Zhu Y, Qiu P, Ji Y. TCGA-assembler: open-source software for retrieving and processing TCGA data. Nat Methods 2014;11:599-600.
17. Davis S, Meltzer PS. GEOquery: a bridge between the Gene Expression Omnibus (GEO) and BioConductor. Bioinformatics 2007;23:1846-7.
18. Director's Challenge Consortium for the Molecular Classification of Lung Adenocarcinoma; Shedden K, Taylor JM, et al. Gene expression-based survival prediction in lung adenocarcinoma: a multi-site, blinded validation study. Nat Med 2008;14:822-7.
19. Stuart T, Butler A, Hoffman P, et al. Comprehensive Integration of Single-Cell Data. Cell 2019;177:1888-1902.e21.
20. Wilkerson MD, Hayes DN. ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. Bioinformatics 2010;26:1572-3.
21. Linderman GC, Steinerberger S. Clustering with t-SNE, provably. SIAM J Math Data Sci 2019;1:313-32.
22. Diaz-Papkovich A, Anderson-Trocmé L, Gravel S. A review of UMAP in population genetics. J Hum Genet 2021;66:85-91.
23. Anders S, Huber W. Differential expression analysis for sequence count data. Genome Biol 2010;11:R106.
24. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 2014;15:550.
25. Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A 2005;102:15545-50.
26. Yu G, Wang LG, Han Y, et al. clusterProfiler: an R package for comparing biological themes among gene

clusters. OMICS 2012;16:284-7.

27. Barbie DA, Tamayo P, Boehm JS, et al. Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1. Nature 2009;462:108-12.
28. Mayakonda A, Lin DC, Assenov Y, et al. Maftools: efficient and comprehensive analysis of somatic variants in cancer. Genome Res 2018;28:1747-56.
29. Chan BKC. Data Analysis Using R Programming. Adv Exp Med Biol 2018;1082:47-122.
30. Decaluwé H, De Leyn P, Vansteenkiste J, et al. Surgical multimodality treatment for baseline resectable stage IIIA-N2 non-small cell lung cancer. Degree of mediastinal lymph node involvement and impact on survival. Eur J Cardiothorac Surg 2009;36:433-9.
31. Yun JK, Lee GD, Choi S, et al. Comparison between lymph node station- and zone-based classification for the future revision of node descriptors proposed by the International Association for the Study of Lung Cancer in surgically resected patients with non-small-cell lung cancer. Eur J Cardiothorac Surg 2019;56:849-57.
32. Wang L, Zhan C, Gu J, et al. Role of Skip Mediastinal Lymph Node Metastasis for Patients With Resectable Non-small-cell Lung Cancer: A Propensity Score Matching Analysis. Clin Lung Cancer 2019;20:e346-55.
33. Park BJ, Kim TH, Shin S, et al. Recommended Change in the N Descriptor Proposed by the International Association for the Study of Lung Cancer: A Validation Study. J Thorac Oncol 2019;14:1962-9.
34. Liu Z, Liang H, Lin J, et al. The incidence of lymph node metastasis in patients with different oncogenic driver mutations among T1 non-small-cell lung cancer. Lung Cancer 2019;134:218-24.
35. Deng C, Zhang Y, Ma Z, et al. Prognostic value of epidermal growth factor receptor gene mutation in resected lung adenocarcinoma. J Thorac Cardiovasc Surg 2021;162:664-674.e7.
36. Wang H, Sun M, Guo J, et al. 3-O-(Z)-coumaroyloleanolic acid overcomes Cks1b-induced chemoresistance in lung cancer by inhibiting Hsp90 and MEK pathways. Biochem Pharmacol 2017;135:35-49.
37. Zhao H, Iqbal NJ, Sukrithan V, et al. Targeted Inhibition of the E3 Ligase SCF(Skp2/Cks1) Has Antitumor Activity in RB1-Deficient Human and Mouse Small-Cell Lung Cancer. Cancer Res 2020;80:2355-67.
38. Chen C, Peng S, Li P, et al. High expression of NEK2 promotes lung cancer progression and drug resistance and is regulated by mutant EGFR. Mol Cell Biochem 2020;475:15-25.

Cite this article as: Zhu J, Wang W, Ma Y, Zhao J, Xiong Y. Development and validation of a gene-based classification model for pN2 lung adenocarcinoma. Transl Lung Cancer Res 2023;12(3):494-509. doi: 10.21037/tlcr-23-16

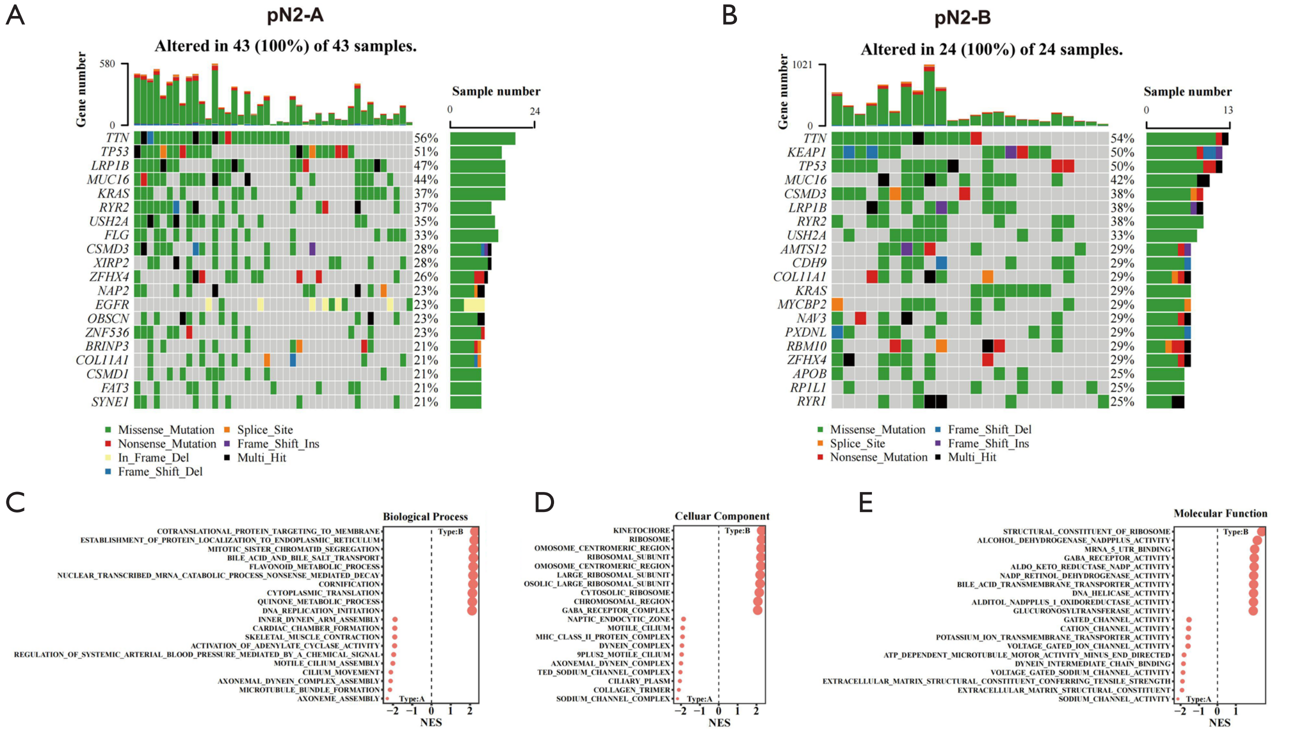

Figure S1 Biological significance and drive gene mutation status of the pN2-A/B molecular types in TCGA-LUAD. (A) pN2-A and (B) pN2-B mutation status of the top 20 genes. (C) GO-biological process, (D) GO-cellular component, and (E) GO-molecular function.

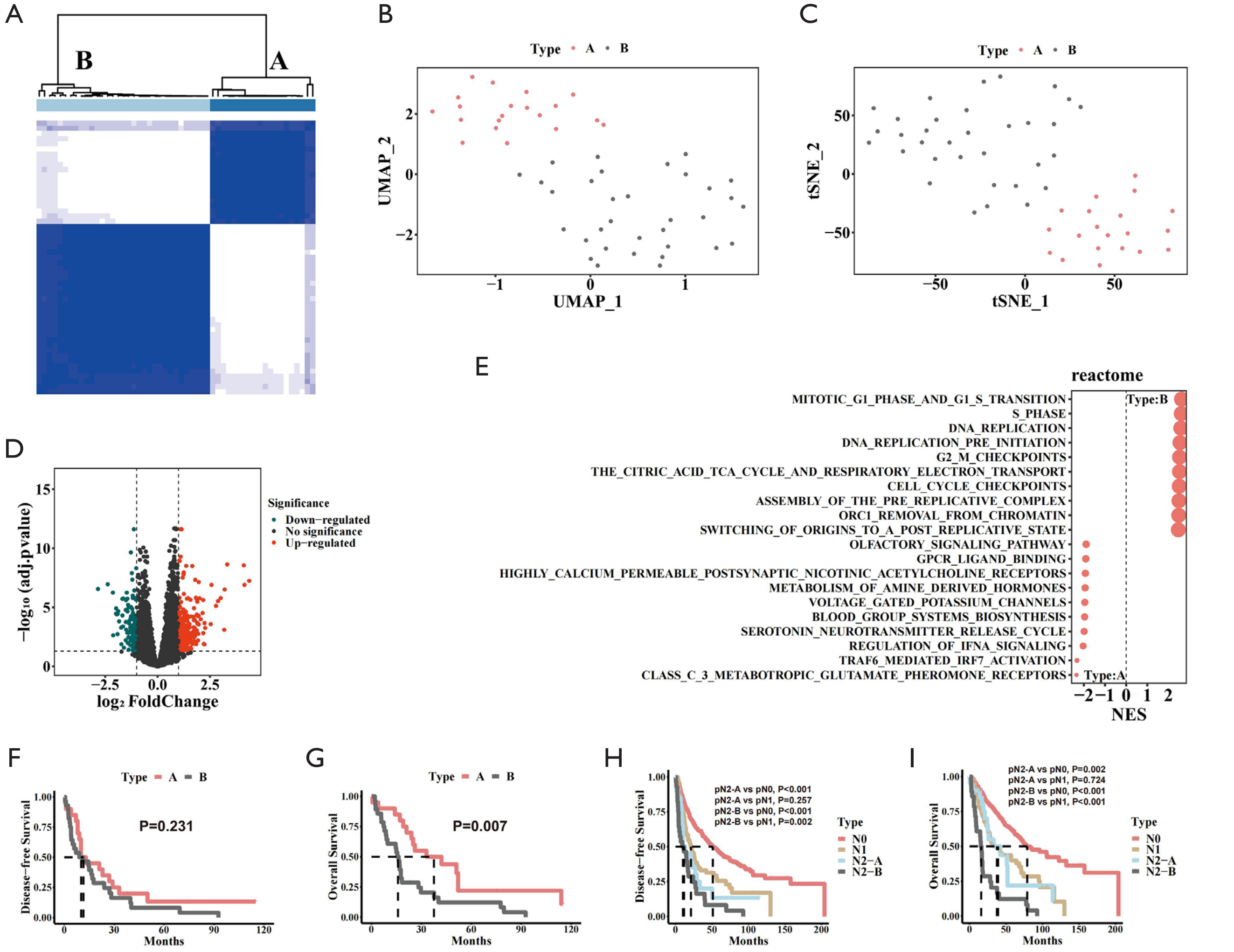

Figure S2 Validation of the molecular typing of LUAD with pN2 metastasis and the clinical and biological assessments using the GSE68465 dataset. (A) Consensus clustering of 53 patients into the pN2-A and pN2-B types based on the gene expression profile. (B) U-MAP and (C) t-SNE were applied to evaluate the effects of classification into the pN2-A/B molecular types, the red and gray circles represent the gene expression profile of each patient with the two subtypes of pN2-A and pN2-B, respectively. (D) Volcano plot showing the gene expression differences between pN2-B and pN2-A. (E) Reactome Pathway Database analysis was performed to assess biological enrichment in the pN2-A/B types by GSEA. Comparisons of the (F) disease-free survival and (G) overall survival among the pN2-A/B molecular groups. Comparisons of the (H) disease-free survival and (I) overall survival and the Kaplan-Meier curves for pN0, pN1, pN2-A, and pN2-B LUAD.

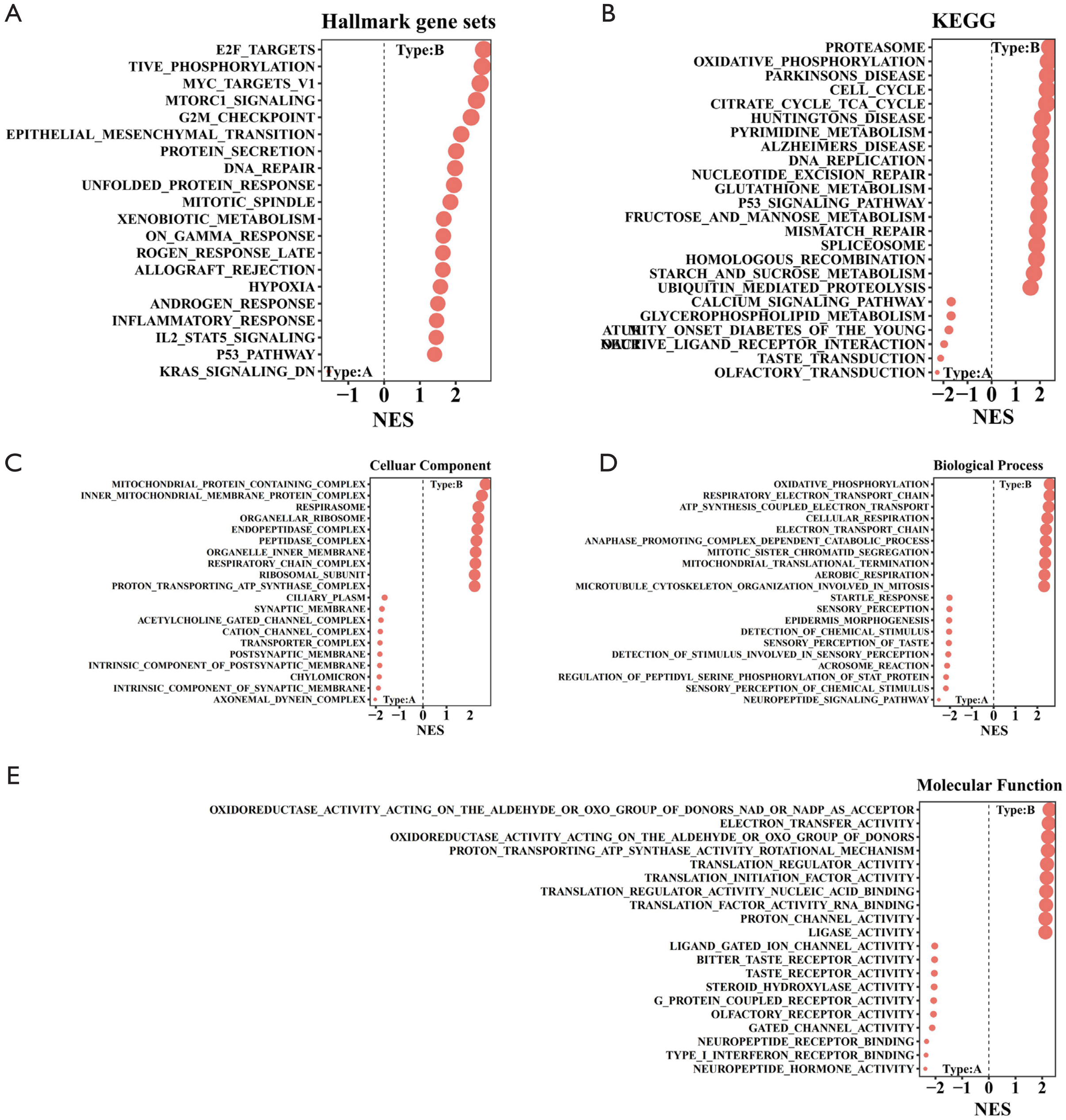

Figure S3 Validation of the biological significance of pN2-A/B molecular types in the GSE68465 dataset. (A) Hallmark, (B) KEGG analyses of signaling pathways. (C) GO-Cellular Component, (D) GO-Biological Process, and GO-Molecular Function (E) analyses were performed to assess biological enrichment in the pN2-A/B types by GSEA.

https://dx.doi.org/10.21037/tlcr-23-16

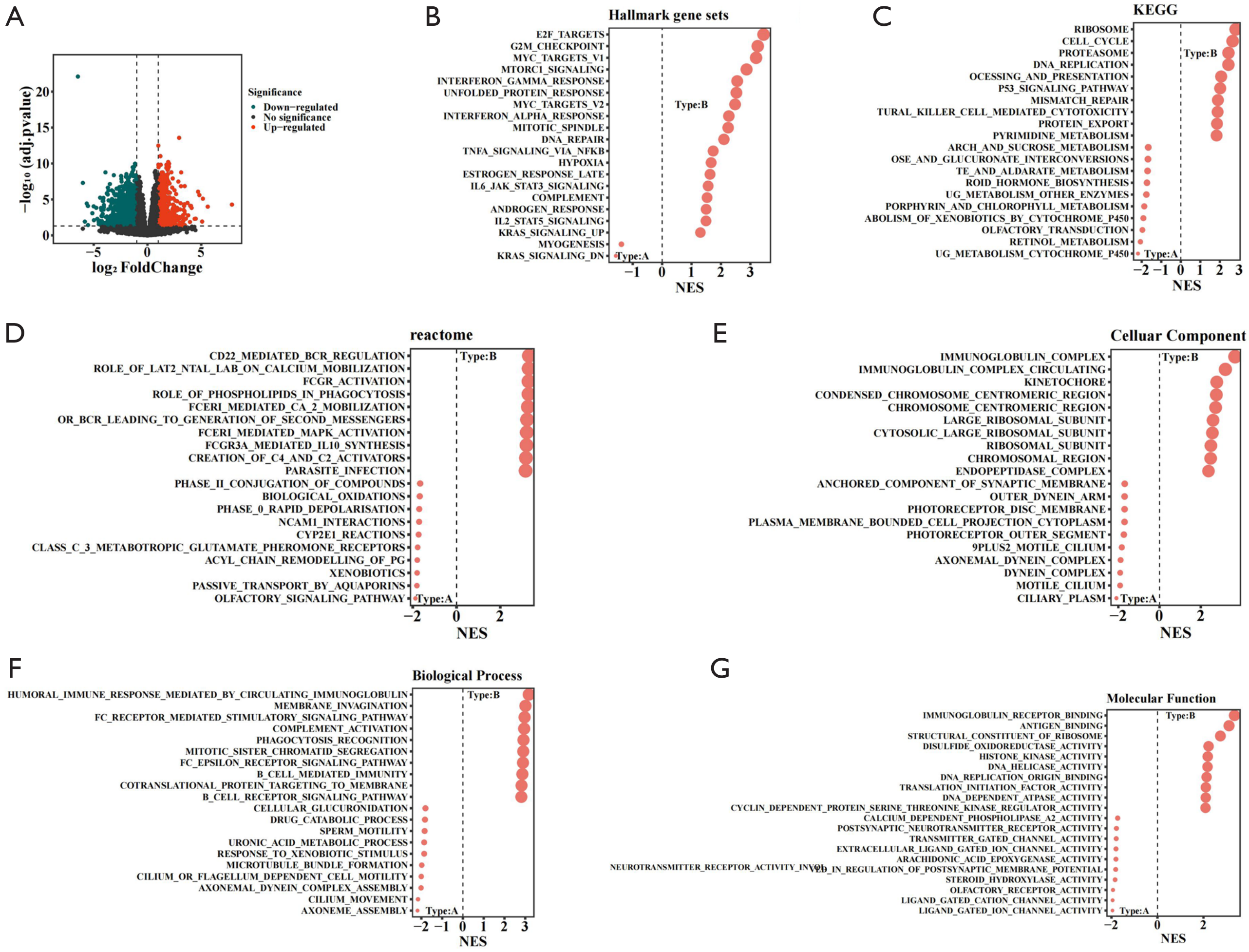

Figure S4 The biological significance of the pN2-A/B molecular types in our dataset. (A) Volcano plot showing gene expression differences between the pN2-B and pN2-A groups. (B) Hallmark and KEGG Pathway (C) analyses were performed to assess the biological enrichment in the pN2-A/B types. (D) Reactome Pathway Database analysis was conducted to assess biological enrichment in the pN2-A/B types by GSEA. (E) GO-Cellular Component, GO-Biological Process (F), and GO-Molecular Function (G) analyses were carried out to assess biological enrichment in the pN2-A/B types by GSEA.