Original Article

Identification of a novel prognostic signature based on vitamin metabolism clustering-related genes in lung adenocarcinoma

Yu Chen[1,2#], Yupeng Jiang[1#^], Xionghui Li[3], Hong Huang[4], Yangying Zhou[2], Chenzi Zhang[5], Shunjun Wang[6], Hanibal Bohnenberger[7], Yawen Gao[1]

[1]Department of Oncology, The Second Xiangya Hospital, Central South University, Changsha, China; [2]Department of Oncology, Xiangya Hospital, Central South University, Changsha, China; [3]Department of Critical Medicine, Hunan Academy of Traditional Chinese Medicine Affiliated Hospital, Changsha, China; [4]Department of Clinical Medicine, Guilin Medical University, Guilin, China; [5]Department of Hematology, Xiangya Hospital, Central South University, Changsha, China; [6]Department of Chest Surgery, Qinghai Red Cross Hospital, Xining, China; [7]Institute of Pathology, University Medical Center, Göttingen, Germany

Contributions: (I) Conception and design: Y Jiang, Y Gao; (II) Administrative support: Y Jiang, Y Gao; (III) Provision of study materials or patients: X Li, Y Zhou; (IV) Collection and assembly of data: Y Chen, H Huang; (V) Data analysis and interpretation: Y Zhou, H Bohnenberger, C Zhang; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

[#]These authors contributed equally to this work.

Correspondence to: Yawen Gao, MD. Department of Oncology, The Second Xiangya Hospital, Central South University, No. 139 Renmin Road, Changsha 410000, China. Email: ygao6@csu.edu.cn.

Background: Vitamins, and their metabolic processes play essential regulatory roles in controlling proliferation, differentiation, and growth in carcinogenesis. However, the role of vitamin metabolism in lung adenocarcinoma (LUAD) has rarely been reported. Here, we established a novel prognostic model based on vitamin metabolism-related genes in LUAD.

Methods: In this research, we aimed to identify vitamin metabolism associated with differentially expressed genes (DEGs) in LUAD utilizing The Cancer Genome Atlas (TCGA)-LUAD, GSE68465 and GSE72094 data. Unsupervised clustering classified patients into distinct subgroups. By utilizing least absolute shrinkage and selection operator (LASSO)-Cox regression analysis, vitamin metabolism-related genes could be used to construct prognostic model. Then the vitamin metabolism gene-related risk score (VRS) was calculated based on best cut-off splitting. Kaplan-Meier analysis, time-dependent receiver operating characteristic (ROC) analysis, univariate and multivariate Cox analyses, chemotherapeutic drugs sensitivity analysis, immune infiltration analysis and nomogram were conducted to verify our models' accuracy. Finally, CPS1 was identified as a relevant diagnostic marker using Random Forests algorithms, single-cell RNA sequencing data was used to confirm its expression.

Results: We investigated the relationship between vitamin metabolism patterns, overall survival (OS), and immune infiltration levels of patients with LUAD. A prognostic signature consisting of 11 genes was developed, which was able to classify patients into high and low VRS groups. Through gene enrichment analysis, cell cycle was mainly enriched. Compared to the low VRS group, the high VRS group exhibited poorer OS, as demonstrated by the Kaplan-Meier survival analysis. Furthermore, VRS was identified as an independent predictor of poor prognosis and poor OS, as indicated by both univariate and multivariate Cox regression analyses. Additionally, a nomogram was constructed to improve the accuracy of survival predictions in LUAD patients. We also found that the two groups of patients might respond differently to immune targets and anti-tumor drugs. CPS1 was identified as a relevant diagnostic marker and the expression was also as confirmed by single-cell RNA sequencing data.

Conclusions: Overall, our findings suggest that vitamin metabolism can influence the prognosis of LUAD patients, and our prognostic signature represents a potentially helpful resource for predicting patient

^ ORCID: 0000-0001-8971-2079.

outcomes and informing clinical decision-making.

Keywords: Vitamin metabolism; lung adenocarcinoma (LUAD); signature; prognosis

Submitted Mar 13, 2024. Accepted for publication May 09, 2024. Published online May 29, 2024.
doi: 10.21037/tlcr-24-245
View this article at: https://dx.doi.org/10.21037/tlcr-24-245

Introduction

Globally, lung cancer is one of the most common types of carcinoma and causes the highest number of cancer-related mortality, accounting for approximately 12.17% of newly diagnosed cancers and 19.27% of cancer related deaths (1). Lung adenocarcinoma (LUAD) has emerged as the primary type of non-small cell lung cancer (NSCLC) in the pathological classification (2). Despite there are significant advancements in cancer treatment technologies, including surgery, radiotherapy, targeted therapy, and immunotherapy, the prognosis for patients with LUAD remains poor, especially for those diagnosed at an advanced stage, resulting in a less than 20% 5-year survival rate (3). Furthermore, despite patients being classified in the same stage, variation has been observed in their prognoses (4). An ability to distinguish the risk of LUAD patients and personalized therapy may help to improve the prognosis of LUAD (5). Therefore, there is a pressing need to identify a more precise and personalized prognostic model to predict patient outcomes and guide treatment decisions.

Vitamins act as indispensable organic substances that contribute to normal physiological metabolic function of the body, such as maintaining immune function, promoting cell growth, mediating cell signaling, and repairing DNA (6,7). Recently, numerous studies have identified an association between the progression of lung cancer and vitamin metabolism. For instance, vitamin D, is a widely recognized fat-soluble vitamin that plays a critical function in promoting tumor cell apoptosis and inhibiting their angiogenesis via its ability to modulate gene expression and cellular signaling pathways related to the vitamin D receptor (VDR) (8-11). Deficiencies in vitamin C and vitamin E can hinder the immune function and increase the risk of cancer (12-15). Moreover, it is well known that folate and vitamin B12 are essential vitamins in synthesis of purine and pyrimidine nucleotides, and pre-treatment with these vitamins can optimize the availability of pemetrexed, reducing adverse effects and enhancing the therapeutic efficacy of LUAD patients (16). However, few studies have concentrated on constructing diagnostic and prognostic models based on vitamin utilization related genes; additional investigations are needed.

In this study, in order to explore the influence of vitamin metabolism on the survival and its implications for clinical decision-making, we established a vitamin metabolism model and made statistical analysis of the differentially expressed genes (DEGs). External validation was exhibited using two independent Gene Expression Omnibus (GEO) datasets (GSE68465 and GSE72904). Our findings offer valuable insights into the vitamin-metabolic crosstalk and its potential for predicting prognosis and facilitating treatment strategies in LUAD. We present this article in accordance with the TRIPOD reporting checklist (available at https://tlcr.amegroups.com/article/view/10.21037/tlcr-24-245/rc).

Highlight box

Key findings

- Novel prognostic signature: vitamin metabolism's impact on lung adenocarcinoma (LUAD) survival and clinical decision-making.

What is known and what is new?

- Previous studies have linked lung cancer progression to vitamin metabolism, including the roles of vitamin D, vitamin C, vitamin E, folate, and vitamin B12.
- We introduce a novel prognostic model based on vitamin metabolism-related genes in LUAD, revealing specific genes associated with prognosis, immunotherapeutic responses, and drug sensitivity, providing a foundation for personalized therapy and emphasizing the importance of considering vitamin metabolism in LUAD treatment.

What is the implication, and what should change now?

- This study urges a transformative shift in LUAD treatment, emphasizing the integration of vitamin metabolism considerations for personalized prognosis assessments and tailored therapeutic strategies.
- Further research is essential to validate the multi-parameter prognostic model through large-scale real-world validation and experimental verification.

Methods

Data acquisition and differential analysis

RNA sequencing data and relevant clinical outcomes pertaining to LUAD patients were acquired from The Cancer Genome Atlas (TCGA) database, including normal tissues and LUAD tissues. The validation cohorts consisted of the GSE72094 and GSE68465 which were obtained from the GEO database, along with their corresponding clinical information (17,18). The raw data were performed in transcripts per million (TPM) format and log-transformed. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013).

Selection of vitamin metabolism-related genes

Altogether, 111 vitamin metabolism-related genes (GOBP_VITAMIN_METABOLIC_PROCESS.v2022.1) were identified from the Molecular Signature Database (MsigDB; https://www.gsea-msigdb.org/) to perform DEGs analysis between LUAD and normal tissues. A total of 19 vitamin metabolism-related genes were selected with absolute fold change [$|\log_2$ fold change (FC)$|$] >1.0 and adjusted P value <0.05 using the "limma" R package.

Unsupervised consensus clustering

The "ConsensusClusterPlus" R package was utilized to investigate the potential impact of different vitamin metabolism patterns on prognosis according to gene expression profiles of LUAD. Furthermore, the optimal cluster number was decided based on the consensus matrix and the cumulative distribution function (CDF) curves.

Development and validation of a prognostic signature

To construct a clinical prognostic signature of vitamin metabolism in LUAD, univariate Cox analysis was conducted to screen out genes based on vitamin metabolism related DEGs in TCGA. Next, the training set consisted of the TCGA-LUAD cohort, and the GSE68465 and GSE72904 cohorts were utilized as external validation sets (19,20). The least absolute shrinkage and selection operator (LASSO)-Cox regression analysis was employed to select genes and minimize redundant genes (21,22). The vitamin metabolism gene-related risk score (VRS) was calculated based on the expression of gene expression values and its corresponding coefficients based on the following equation: risk score = $\sum$ genes Cox coefficient × gene expression. Subsequently, patients were divided into low or high VRS groups using the most suitable cutoff value. To further identify the major constituent of VRS formula, the "randomForest" R package in R was used to perform the random forest technique in this study. In order to examine the predictive performance of our risk model, Kaplan-Meier survival curves and time-dependent receiver operating characteristic (ROC) curves were drawn.

Evaluation of clinical features and subgroup analysis

To better understand the interrelation between prognostic model and clinical outcomes of LUAD, risk scores of several clinical features were analyzed, comprising age, gender, tumor (T) stage, node (N) stage, metastasis (M) stage, pathological TNM stage, and primary outcome. To better assess the prognostic relevance of the signature model, subgroup analysis based on pathological TNM stages, ages, and genders was conducted. The R package "survival" was utilized to perform univariate and multivariate Cox regression analyses based on both clinical information and the risk score.

Development and examination of a prognostic nomogram model

A predictive nomogram was built using VRS and clinical characteristics by "rms" R package. For evaluating the nomogram precision, calibration curves were generated.

Differential analysis of immune cell infiltration and immune checkpoint function

Immune cell infiltration was identified using CIBERSORT algorithms. We utilized the "gsva" R package for single-sample gene set enrichment analysis (ssGSEA) to evaluate the activities of 13 immune pathways. Additionally, immune checkpoint-related genes were obtained following prior methods (23).

scRNA-seq and immunohistochemical (IHC) data extraction

Tumor Immune Single-cell Hub 2 (TISCH2) is an scRNA-seq database accessible at http://tisch.comp-genomics.org/, dedicated to providing a detailed characterization of the tumor microenvironment (TME) at the single-cell

level (24), in this study we used TISCH2 to further analysis GSE148071.

To explore variances between normal and tumor tissues, we obtained IHC data from 10 LUAD cases using the Human Protein Atlas public database available at https://www.proteinatlas.org/.

Drug sensitivity analysis

The "pRRophetic" R package was employed to identify frequently prescribed chemotherapeutic drugs and to calculate its half-maximal inhibitory concentration (IC_{50}). Then, according to the NSCLC National Comprehensive Cancer Network (NCCN; https://www.nccn.org) guidelines, first-line drugs (cisplatin, paclitaxel, docetaxel, gemcitabine, vinblastine, and etoposide) that were used to treat LUAD patients were submitted to drug sensitivity analysis (25). Wilcoxon rank test was employed to compare the variation in IC_{50} values between the two VRS subgroups.

Statistical analysis

Statistical analysis was conducted utilizing the R software package (4.1.0; The R Foundation for Statistical Computing, Vienna, Austria). Wilcoxon symbolic rank test was employed to investigate discrepancies in the composition of immunoinfiltrated cells between the two groups. Kaplan-Meier survival analysis between subgroups was executed using the log-rank test. Statistical significance was set at $P<0.05$.

Results

Investigation, functional analysis, and clustering of vitamin metabolism DEGs

We included 493 LUAD patients in the training sets and 825 LUAD patients in the validation sets for analysis. The expression data of LUAD tissues and normal lung tissues were downloaded and analyzed. To explore the relationship between the potential prognosis of LUAD and vitamin metabolism, a set of 111 DEGs related to vitamin metabolism in LUAD was identified. Among them, 21 DEGs were found to be significantly differentially expressed ($|\log_2 FC| >1$ and $P<0.05$) between tumor site and normal tissues in TCGA dataset. As shown in *Figure 1A*, 10 DEGs (dark red dots) were upregulated in tumor site, and 11 DEGs (dark blue dots) were downregulated.

Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) functional analysis were utilized to further explore the molecular mechanisms. GO analysis demonstrated that mitotic nuclear division and chromosome segregation were significantly enriched in biological processes. According to the KEGG pathway analysis, the DEGs were closely associated with the cell cycle, asthma, P53 signaling pathway, and human T-cell leukemia virus1 infection (*Figure 1B,1C*).

Unsupervised consensus clustering analysis was conducted to distinguish the vitamin metabolism patterns in LUAD (*Figure 1D*). According to the overall survival (OS) of LUAD, a total of 21 DEGs were categorized into distinct clusters, and a considerable difference in OS was identified between the two different clusters (*Figure 1E*). The two GEO datasets were also analyzed by the similar methods (*Figure 1F,1G*). Among the 21 DEGs, a total of 19 genes were identified from the training and validation sets (*Figure 1H*). In addition, unsupervised clustering analysis also revealed the weighted value of clinical characteristics between the two cluster groups (Figure S1A). Immune infiltration levels and the expression of checkpoint inhibitors are also displayed in Figure S1B,S1C.

Construction and validation the prognostic model

The differential expression analysis of vitamin metabolism-related genes between normal and tumor samples in the TCGA-LUAD cohort was performed using the "limma" package. We utilized the TCGA-LUAD cohort as our training set and validated our findings among the GSE68465 and GSE72904 cohorts. To evaluate the prognostic value of vitamin metabolism-related genes, 250 of 396 DEGs were detected between the two clusters (*Figure 2A,2B*). In order to avoid overfitting, these 250 genes were screened out using Cox regression to identify the most important DEGs that are associated with survival. Then, we selected 11 novel prognostic genes, *SFTPB*, *PLK1*, *LYPD3*, *KRT6A*, *ABCC2*, *HMMR*, *PFKP*, *CPS1*, *FOSL1*, *CDKN3*, and *LOXL2*, as variables for the prognostic signature (*Figure 2C,2D*). LUAD samples were categorized into high VRS and low VRS subgroups by using the optimistic cutoff value to build our prediction model. The model formula was utilized to compute risk scores of each sample: risk score = LYPD3 × 0.018 + KRT6A × 0.047 + ABCC2 × 0.048 + HMMR × 0.050 + PFKP × 0.051+ CPS1 × 0.065 + FOSL1 × 0.067 + CDKN3 × 0.095 + LOXL2 × 0.101 – SFTPB × 0.042 – PLK1 × 0.003.

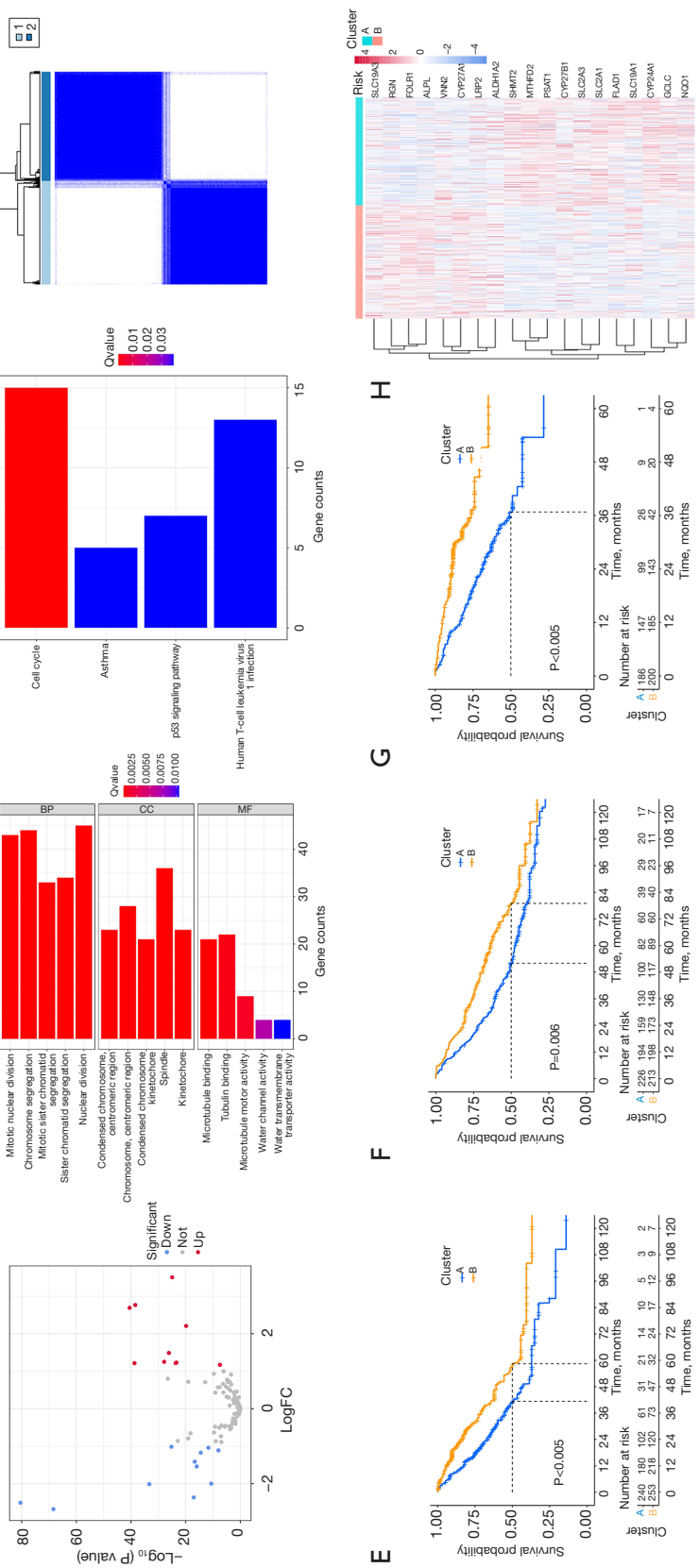

Figure 1 Exploration of vitamin metabolism-related DEGs. (A) 21 DEGs related to vitamin metabolism; (B) GO enrichment and (C) KEGG pathways of vitamin metabolism-related DEGs; (D) heatmap of the consensus matrix; (E-G) Kaplan-Meier curves for the OS of LUAD patients in two clusters; (H) heatmap of unsupervised consensus clustering. FC, fold change; BP, biological process; CC, cellular component; MF, molecular function; DEGs, differentially expressed genes; GO, Gene Ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes; OS, overall survival; LUAD, lung adenocarcinoma.

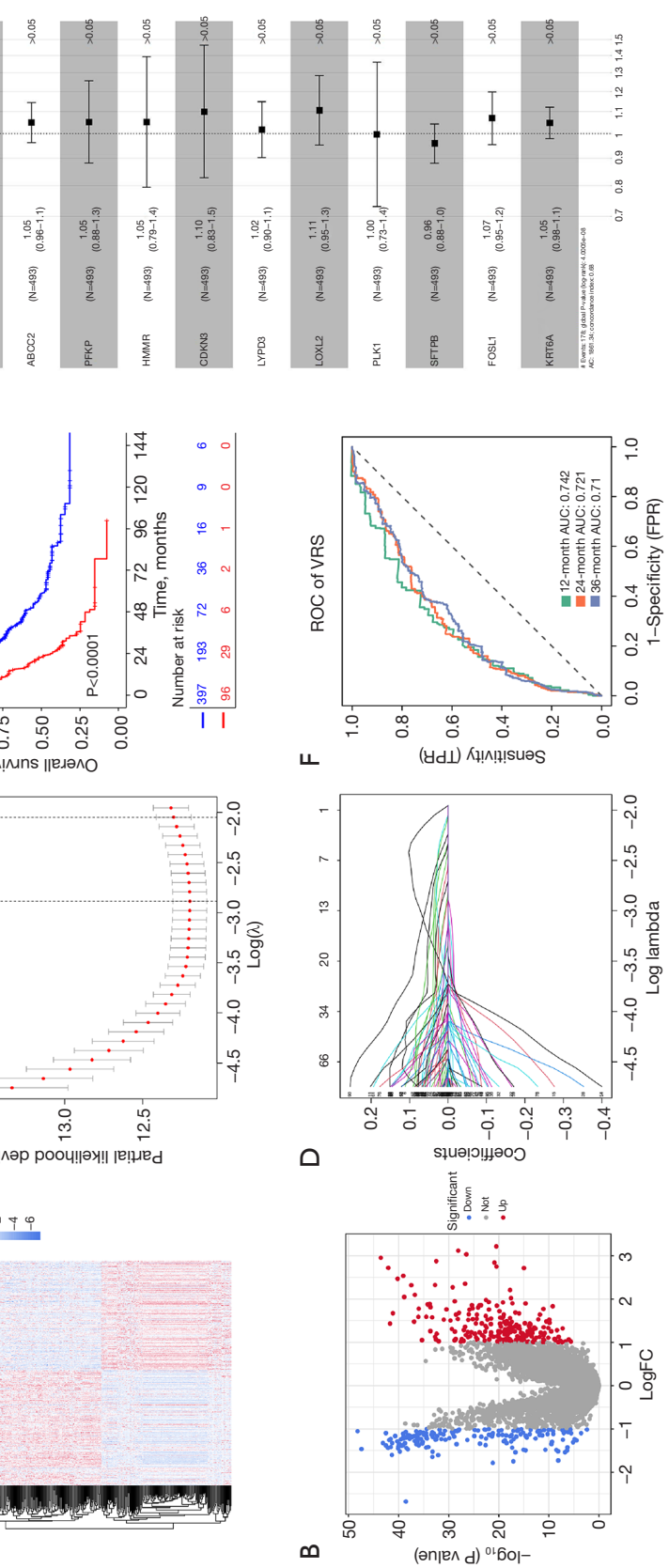

Figure 2 Construction of a vitamin metabolism-related prognostic model. (A) Heatmap and (B) volcano plot of 250 DEGs of unsupervised consensus clustering; (C) application of cross-validation for tuning parameter selection in the LASSO model; (D) the distribution of LASSO coefficients for 11 vitamin metabolism-related DEGs; (E) the Kaplan-Meier survival curve of LUAD with high VRS and low VRS score in training set based on prognostic signature; (F) the ROC curves for the prognostic signature in training set; (G) multivariate Cox regression analysis of DEGs. *, P<0.05. VRS, vitamin metabolism gene-related risk score; FC, fold change; ROC, receiver operating characteristic; TPR, true positive rate; FPR, false positive rate; AUC, area under the curve; CI, confidence interval; AIC, Akaike information criterion; DEGs, differentially expressed genes; LASSO, the least absolute shrinkage and selection operator; LUAD, lung adenocarcinoma.

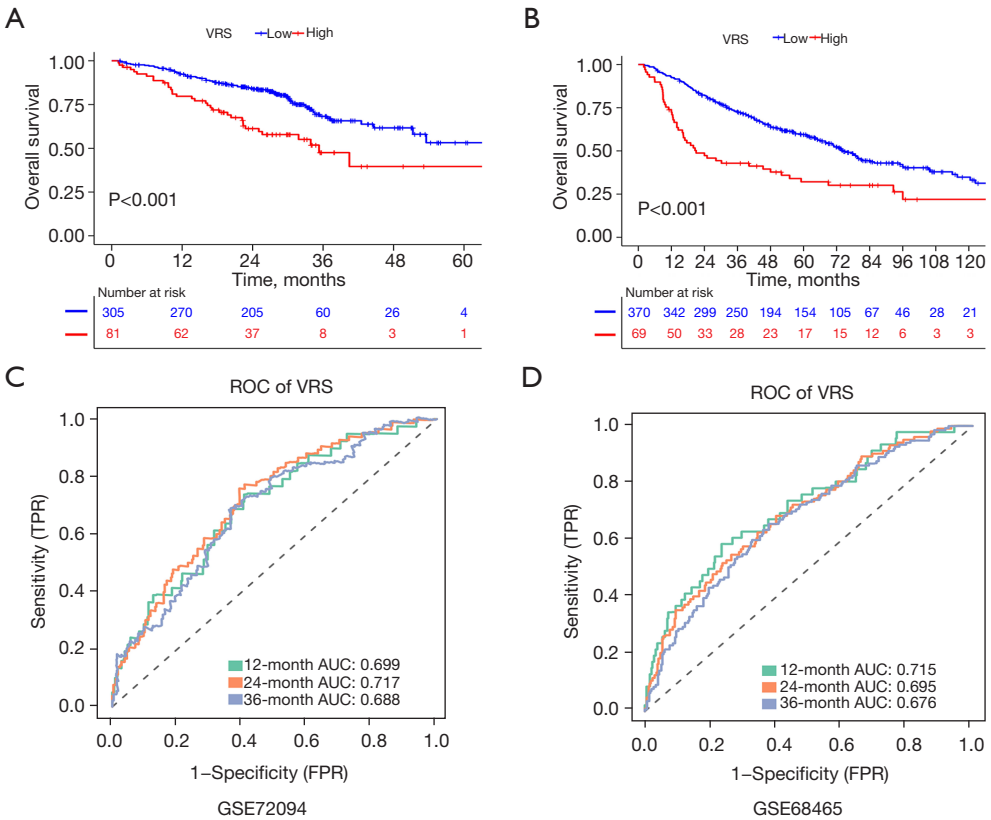

Figure 3 Validation of the vitamin metabolism-related prognostic model in two external validation sets. Kaplan-Meier curves for the OS of LUAD in the high and low VRS groups in the testing sets (A) GSE72094 and (B) GSE68465. ROC curves and their AUC values in the testing sets (C) GSE72094 and (D) GSE68465. VRS, vitamin metabolism gene-related risk score; ROC, receiver operating characteristic; TPR, true positive rate; FPR, false positive rate; AUC, area under the curve; OS, overall survival; LUAD, lung adenocarcinoma.

To assess the prognostic model's predictive performance, we constructed time-dependent ROC curves and Kaplan-Meier survival curves. Compared to the high VRS groups, the Kaplan-Meier analysis showed that the low VRS groups showed a significantly longer OS (*Figure 2E*) (P<0.0001). The areas under the curve (AUCs) of the training cohort for OS at 1, 2, and 3 years were 0.742, 0.721, and 0.71, respectively (*Figure 2F*). The multivariate cox regression analysis of DEGs was shown in *Figure 2G*. The findings obtained from the external validation cohorts were consistent with those observed in the internal validation groups. The patients in validation cohorts (GSE68465 and GSE72904) were stratified into two VRS groups based on the same threshold value used in the TCGA cohort. This analysis revealed that the subgroup of patients with high VRS had a markedly lower OS rate compared to the group with low VRS (*Figure 3A,3B*). For the GSE72094 cohort, the AUCs for predicting 1-, 2-, and 3-year OS were 0.669, 0.717, and 0.688, respectively (P<0.001) (*Figure 3C*). Similarly, for the GSE68465 dataset, the AUCs for predicting 1-, 2-, and 3-year OS were 0.715, 0.695, and 0.676, respectively (P<0.001) (*Figure 3D*).

Relationship between the prognostic model and clinical characteristics

To further investigate the prognostic significance and clinical correlation of LUAD patients, we conducted subgroup analysis based on clinical features revealed significant variations in the VRS levels regarding different age, gender, pathological stages, T stages, N stages, M stages, and pathological stages (*Figure 4*). Patients with low VRS exhibited better outcomes on survival than those with high VRS.

Clinical features and VRS were used to conduct univariate and multivariate Cox analysis (Figure S2A,S2B).

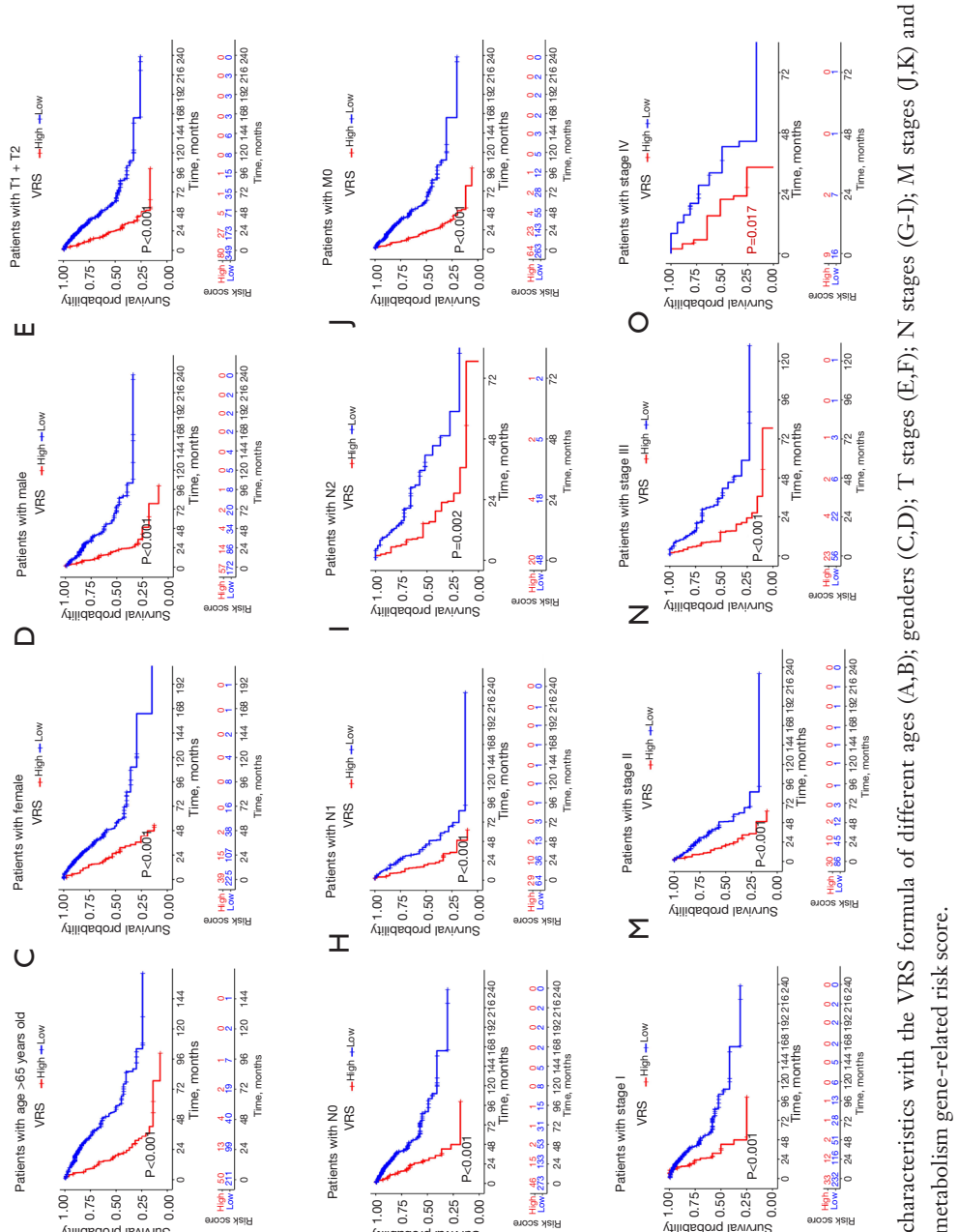

Figure 4 Subgroup analysis of clinical characteristics with the VRS formula of different ages (A,B); genders (C,D); T stages (E,F); N stages (G-I); M stages (J,K) and pathological stages (L–O). VRS, vitamin metabolism gene-related risk score.

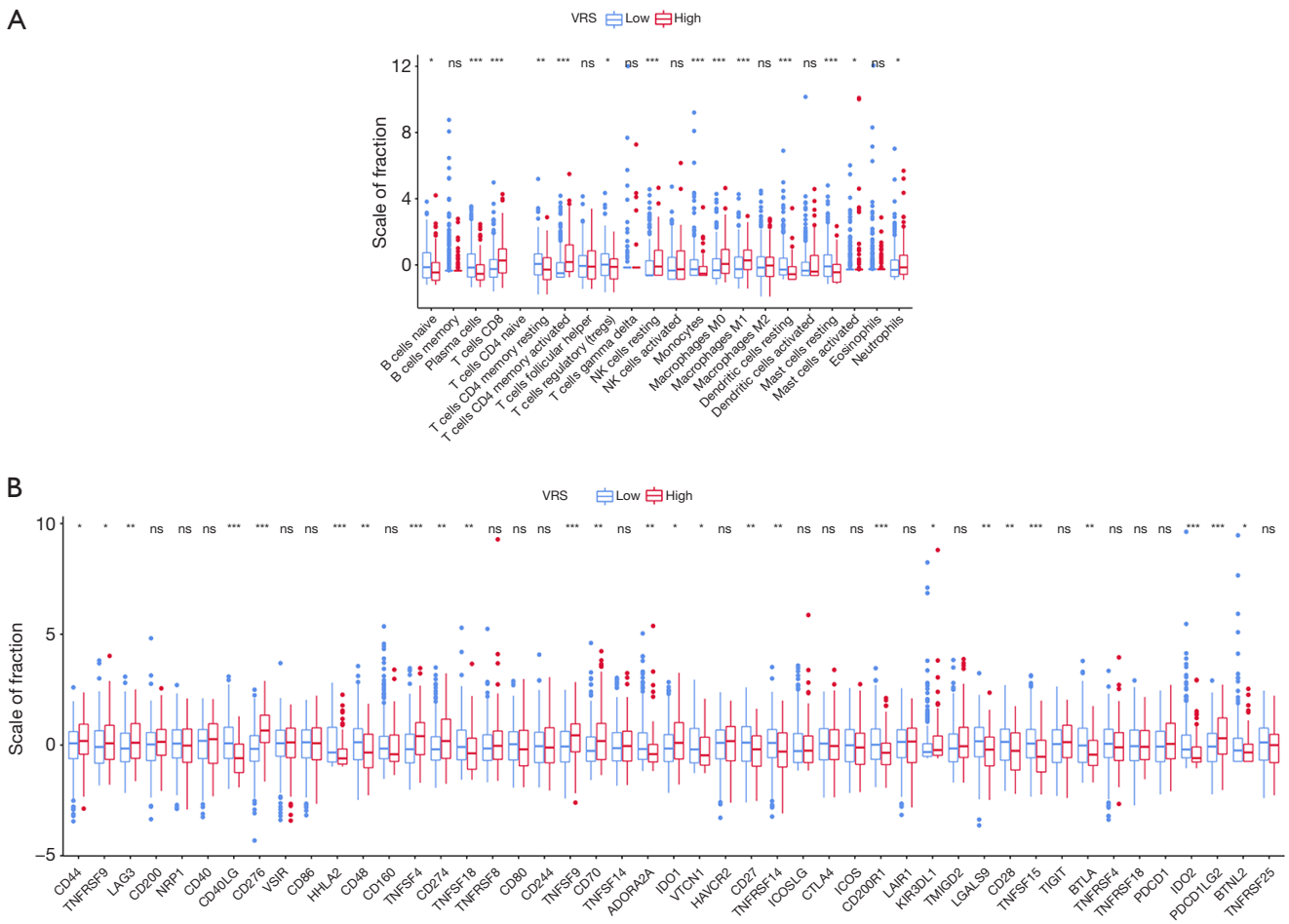

Figure 5 Correlation analysis of the prognostic model and immune cell infiltration (A) and immune checkpoint (B) among high and low VRS groups. *, P<0.05; **, P<0.01; ***, P<0.001; ns, not significant, P>0.05. VRS, vitamin metabolism gene-related risk score; NK, natural killer.

Univariate analysis demonstrated a negative association between OS of LUAD and pathological stage, T stage, N stage, M stage, and VRS. Multivariate analysis further validated independent influence of VRS on the prognosis of LUAD patients.

Correlation between risk signature and immune characteristics

To explore the impact of immune cell infiltration on survival, as shown in *Figure 5A*, our findings revealed immune infiltration levels between different VRS groups, particularly in T cells CD4 memory activated, monocytes, dendritic cells activated, macrophages M0, and mast cells activated. These results suggest a strong association between vitamin metabolism, immune infiltration, and the prognosis of LUAD.

Given the significance of immunotherapy based on checkpoint inhibitors, the immune checkpoints expression was analyzed to investigate potential alterations between the two VRS groups. It was observed there were noteworthy variations in the expression of *CD40*, *CD40LG*, *CD86*, *CD160*, *HHLA2*, *TNFSF4*, *CD244*, *CD200R1*, *TNFSF15*, *IDO2*, and *PDCD1LG2* in the two VRS groups of patients, which indicated different efficacy of immune checkpoint blockade therapy (*Figure 5B*).

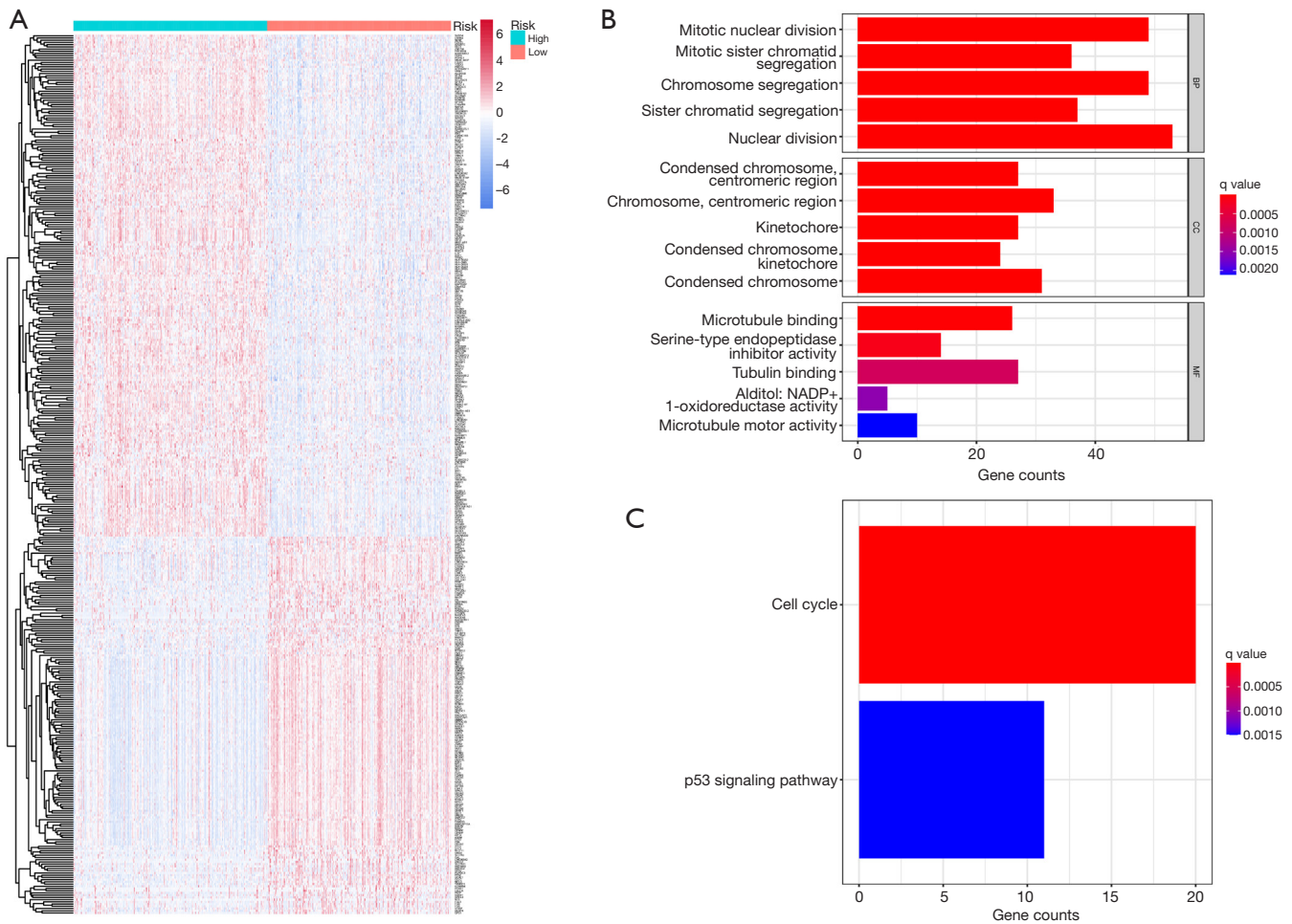

Figure 6 Identification of different VRS groups DEGs and functional enrichment analysis. (A) Heatmap of 591 DEGs between different VRS groups; (B) GO enrichment and (C) KEGG pathways between the high and low VRS patients. NADP, nicotinamide adenine dinucleotide phosphate; BP, biological process; CC, cellular component; MF, molecular function; VRS, vitamin metabolism gene-related risk score; DEGs, differentially expressed genes; GO, Gene Ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes.

Functional enrichment analysis between the two VRS groups

To better understand the fundamental mechanisms that contribute to the variation of the survival, 591 DEGs between different VRS groups were detected (|logFC| >1 and P<0.05) (*Figure 6A*). To gain insights into the functional enrichment of these genes, we employed GO enrichment analysis and KEGG pathway analysis between the two VRS groups. GO analysis revealed that genes involved in mitotic nuclear division were significantly enriched in biological processes. Moreover, the cellular component and molecular function categories were respectively involved in chromosome, centromeric region, and microtubule binding (*Figure 6B*). As shown in *Figure 6C*, KEGG pathway analysis showed that the DEGs related to vitamin metabolism were primarily enriched in the cell cycle pathway. These findings suggest that the DEGs play critical roles in regulating key cellular processes and pathways, emphasizing their potential importance in disease development and progression.

Chemotherapeutic drug sensitivity analysis

For chemotherapy response examination, the "pRRophetic" package was utilized to evaluate the IC_{50} of commonly used chemotherapies. Then, we searched commonly prescribed drugs for treating LUAD patients based on the NCCN guidelines (*Figure 7A*). Among these drugs,

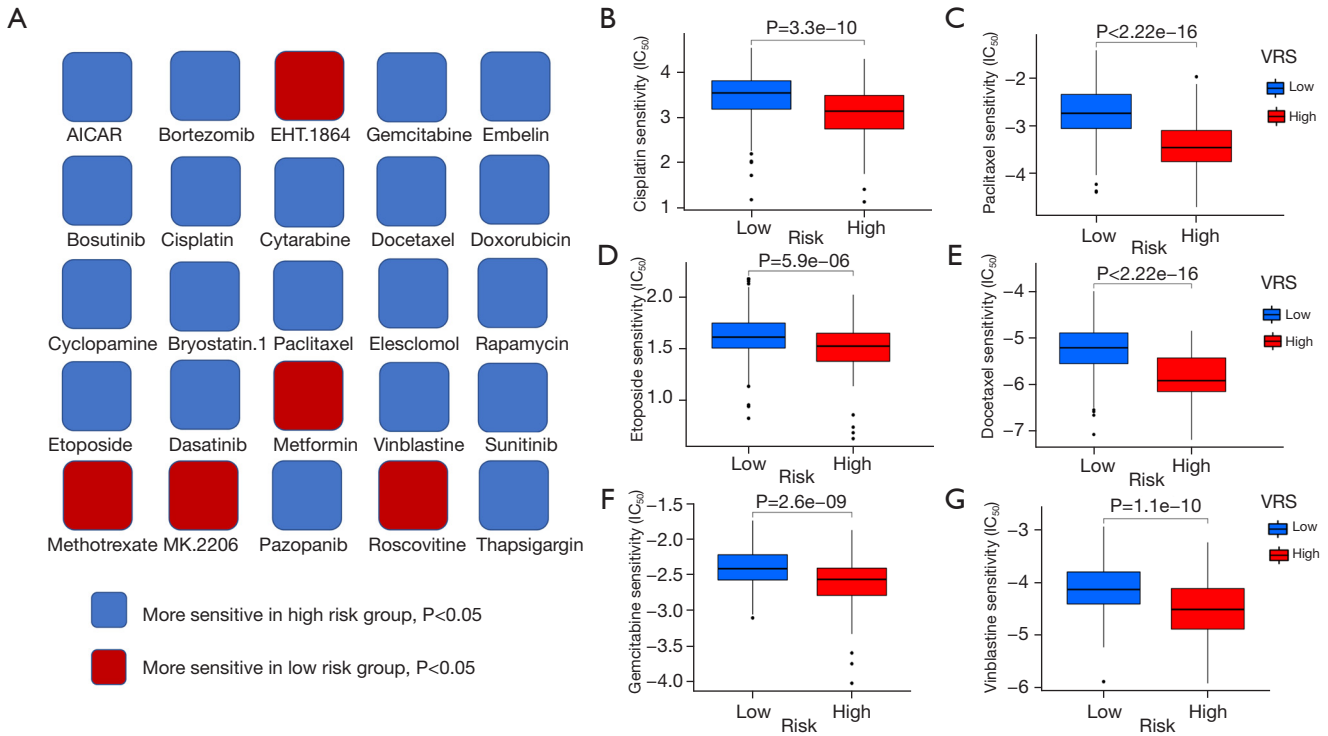

Figure 7 Prediction of chemotherapy susceptibility. (A) Sensitive drugs in high and low VRS groups; (B-G) the IC_{50} of cisplatin, paclitaxel, etoposide, docetaxel, gemcitabine and vinblastine among VRS groups. VRS, vitamin metabolism gene-related risk score; IC_{50}, half-maximal inhibitory concentration.

the different IC_{50} for six common chemotherapy drugs (cisplatin, paclitaxel, docetaxel, gemcitabine, vinblastine, and etoposide) were found to be considerably higher in the low VRS groups (P<0.05). This suggested that LUAD patients categorized into the high VRS group had a greater sensitivity to chemotherapeutic drugs (*Figure 7B-7G*).

Construction and validation of the VRS nomogram

To enhance the practicality of our risk assessment model, an independent-factor-based nomogram was constructed according to the age, gender, stage, and VRS (*Figure 8A*). The reliability and accuracy of the nomogram were well verified through calibration curves, which confirmed its predictive ability for 1-, 2-, and 3-year OS (*Figure 8B*).

Screening diagnostic genes and scRNA-seq extraction

To validate our prognostic model, we utilized Random Forest algorithms to explore the relationship between error rate, the number of classification trees, and the 11 genes ranked by relative relevance in descending order. (*Figure 9A,9B*). Due to its strong correlation, CPS1 was selected as the target gene for further analysis. We analyzed the diagnostic value of CPS1 in LUAD using the ROC curve. As shown in *Figure 9C*, CPS1 may act as a perfect diagnostic marker and increased expression of CPS1 was linked to poor OS (P<0.01). Cybersort analysis was conducted to investigate the impact of CPS1 on immune cell infiltration, revealing a notable inverse correlation between CPS1 expression and the infiltration levels of memory B cells, resting dendritic cells, and resting mast cells (*Figure 9D*).

Subsequently, gene expression data for single-cell RNA sequencing samples were acquired from the GEO database under accession number GSE148071. CPS1 was widely available in malignant, endothelial, and fibroblasts (*Figure 9E,9F*).

Discussion

Despite that there is an increasing number of available

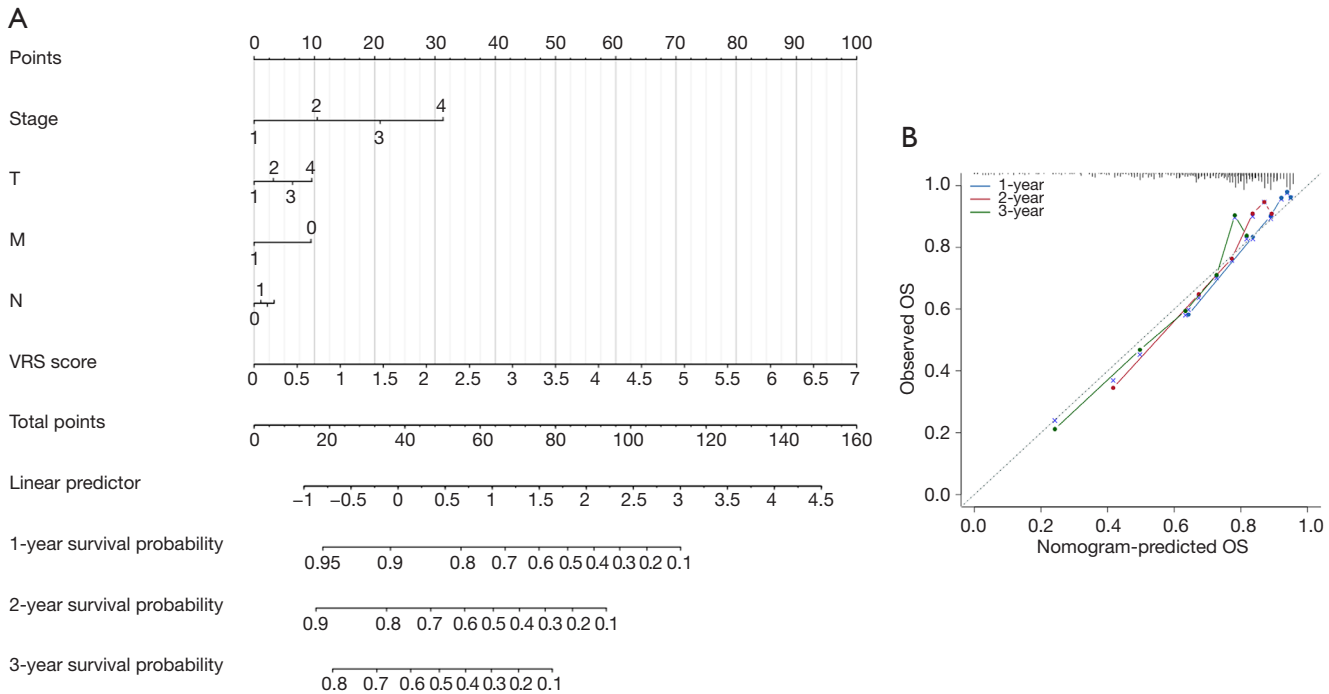

Figure 8 Construction of a prognostic nomogram. (A) A nomogram was developed to predict the 1-, 2-, and 3-year OS of LUAD; (B) calibration curves for predicting OS at 1, 2, and 3 years. VRS, vitamin metabolism gene-related risk score; OS, overall survival; LUAD, lung adenocarcinoma.

treatments, LUAD remains the primary reason for cancer related death, as indicated by the latest epidemiological data (26). Recently, a study has highlighted the significance of vitamins, such as vitamin C, vitamin D, and vitamin A, in the development of LUAD via affecting the TME (27). These vitamins have been shown to not only impact tumor cells proliferation, differentiation, and apoptosis, but also to regulate the function of immune cells (28-30). However, limited research has focused on constructing prognostic models using vitamin-related features. Hence, we propose that there exists a potent relationship between vitamin metabolism and the prognosis of LUAD. This study integrated 19 co-expressed DEGs between LUAD and adjacent normal tissues in TCGA, GSE72094, and GSE68465. Then, two clusters with a significant difference in OS were discovered by using unsupervised consensus clustering analysis. Between the two clusters, 396 DEGs were detected, and finally, 11 genes were selected to establish a precise and dependable prognostic model.

Among these 11 genes, *PLK1*, *CDKN3*, and *FOSL1* are directly involved in regulating the cell cycle. Of the remaining genes, *SFTPB*, *LYPD3*, *HMMR*, and *LOXL2*, have functions of regulating cell proliferation and differentiation; *ABCC2*, *CPS1*, *KRT6A*, and *PFKP* participate in molecules transport and cellular metabolism. These findings suggest that the regulation of vitamin metabolism plays a significant impact in the construction of the prognostic model by modulating the cell cycle.

Furthermore, out of the 11 genes examined, only *SFTPB* and *PLK1* exhibit a negative association with both the VRS and the OS, whereas the other 9 genes present as oncogenes. Among them, *PLK1*, *CDKN3*, and *FOSL1* are most associated with cell cycle predicted by KEGG analysis. *PLK1*, a serine/threonine kinase, participates in multiple stages of mitotic progression, including G2-M transition, spindle formation, chromosome segregation, and cytokinesis (31-33). Besides, a recent study provided evidence to suggest that the upregulation of *PLK1* may contribute to the activation of STAT3, ultimately leading to increased cell proliferation and apoptosis resistance in esophageal cancer (33). *CDKN3*, also known as cyclin-dependent kinase (CDK) inhibitor 3, which belongs to the dual specificity protein phosphatase family, controls normal mitosis and G1/S transition through the CDC2 signaling axis (34,35). The varying expression degrees of *CDKN3*

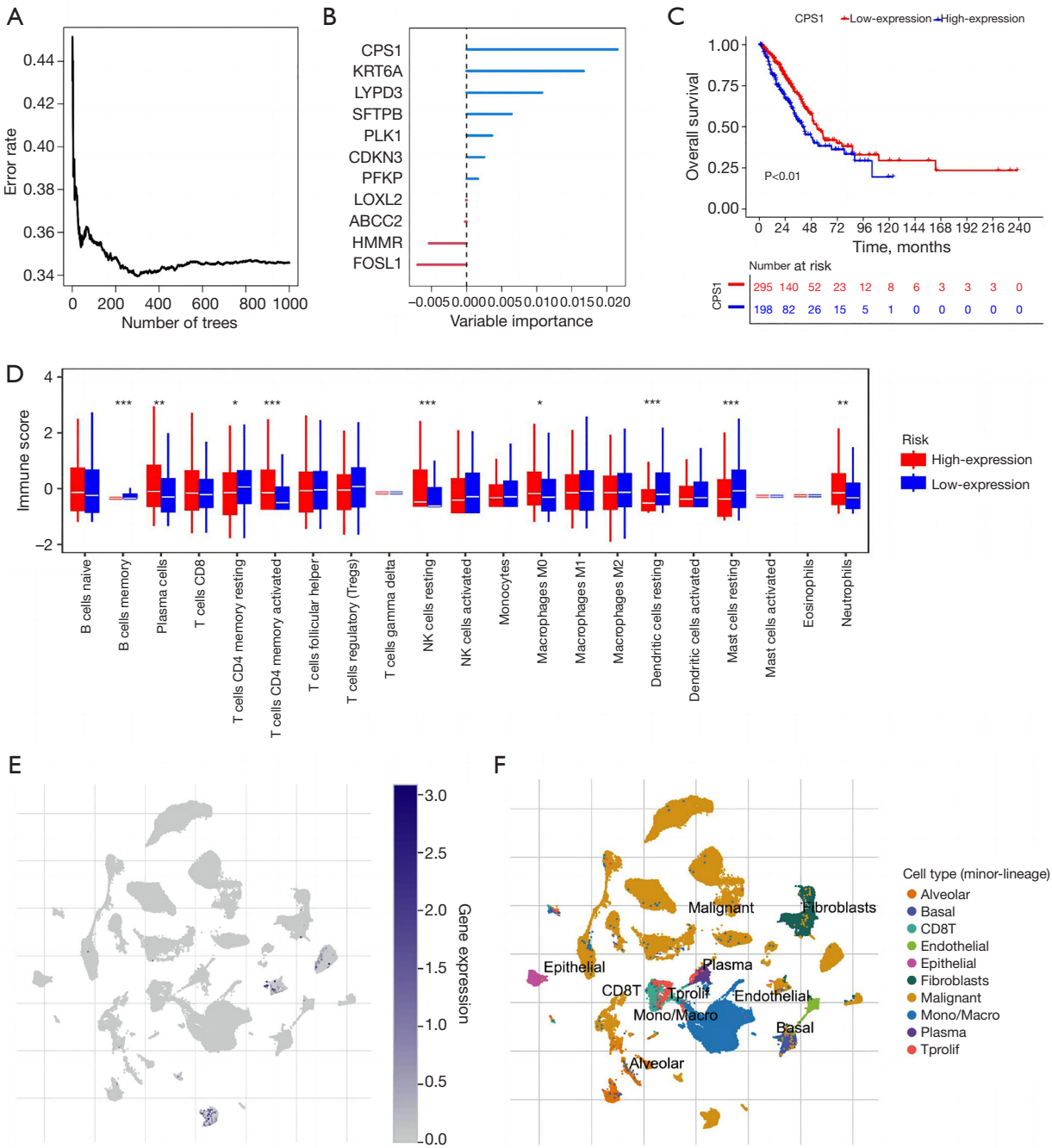

Figure 9 Detection of CPS1, its clinical characteristics and scRNA-seq data extraction. (A,B) Based on random forests algorithm to screen biomarkers; (C) Kaplan-Meier curves for the OS of CPS1 in LUAD; (D) the immune cell infiltration of different CPS1 expression in LUAD; (E,F) basic data and scRNA-seq data of CPS1 in GSE14807. *, P<0.05; **, P<0.01; ***, P<0.001. NK, natural killer; OS, overall survival; LUAD, lung adenocarcinoma; scRNA-seq, single cell RNA sequencing.

exhibit distinct functions across various tumor types. Overexpression of *CDKN3* has been shown to enhance cell proliferation, invasion, apoptosis resistance, and xenograft tumor formation in renal cells (36). In both prostate cancer and breast cancer, researchers have found that depletion of *CDKN3* could result in a reduction in the population of S-phase cells and inhibition CDK2 kinase activity in tumor cells (37,38). Zang *et al.* suggested that *CDKN3* could act as a predictor of prognosis for both LUAD and squamous cell carcinoma, as indicated by their meta-analysis (39). *FOSL1* (FOS-like antigen 1) is a member of the Fos family of transcription factors which can lead to cell cycle arrest and apoptosis in cancer cells (40). Researchers have demonstrated that the overexpression of miR-195-5p can downregulate *FOSL1*, leading to the inhibition of cell cycle entry and an increase in apoptosis (41).

Furthermore, for validation, the AUC values of the prognostic signature were 0.742, 0.721, and 0.71 at 1, 2, and 3 years, respectively. The high VRS groups exhibited a shorter OS compared with the low VRS groups. In the external validation sets (GSE68465 and GSE72904), there were similar results, demonstrating the prognostic model's high practicality and accuracy. After conducting a comprehensive analysis of gene profiles related to vitamin metabolism in samples of LUAD, two distinct subgroups were identified based on the degree of VRS, namely high VRS and low VRS. The results of subgroup analysis indicated a significant association between the VRS and disease stage, suggesting a strong predictive ability of the model in each group. Furthermore, the predictive ability of the model was validated by conducting univariate and multivariate analyses.

Enrichment analysis using GO annotation and KEGG pathways discovered that DEGs were primarily correlated with cell cycle regulating. The cell cycle is a highly regulated process that governs the growth, division, and proliferation of cells (42). Alterations in cell cycle can result in the development and progression of cancer, especially in lung cancer (43,44). For example, overexpression of specific CDKs, such as CDK4/6, can lead to the phosphorylation of retinoblastoma protein and the release of E2F transcription factor, allowing the cell to transition from G1 to S phase, which eventually promotes cell cycle progression and proliferation in LUAD (45-47). Similarly, the stimulation of various signaling pathways, including the MAPK pathways and PI3K/AKT/mTOR, ultimately promotes the progression of cell cycle through the G1 to S phase, allowing DNA replication and cell division by activating CDKs, including CDK1 and CDK4/6 (48,49). This can lead to errors in chromosome alignment during cell division, contributing to the development and progression of lung cancer (50). Therefore, we hypothesized that the variations in survival among LUAD may be associated with vitamin metabolism and cell cycle. Based on the above hypothesis, we further utilized the model to predict clinical prognosis and guide LUAD treatments.

In order to gain deeper insights into the immune-related mechanisms underlying the signature used to predict the prognosis of LUAD patients, the immune infiltration levels were conducted to assess the relationship between different VRS levels (51). The results showed active correlations between the VRS and several anti-tumor immune effector cells, particularly T cells CD4 memory activated, monocytes, dendritic cells activated, macrophages M0, and mast cells activated, suggested that patients with LUAD exhibiting high VRS levels displayed increased responsiveness to immune checkpoint inhibitors. The study also evaluated the sensitivity of chemotherapy in each subgroup patient. Using the "pRRophetic" package, the most commonly used drugs were identified, and then the estimated IC_{50} for six anti-tumor prescriptions (cisplatin, paclitaxel, docetaxel, gemcitabine, vinblastine, and etoposide) were determined based on the NCCN guidelines for NSCLC. The results indicated that patients categorized as the high VRS subgroup exhibited a higher susceptibility to these initial treatment drugs, probably because high-value risk tumor cells may be more susceptible to chemotherapy. The results further highlight the importance of individualized therapy based on VRS classification.

To further validate our hypothesis, random Forest algorithm was used to assess the relevance of 11 genes and CPS1 was selected as a diagnostic marker because of its high correlation. The expression of CPS1 is correlated with poor OS, and has a negative correlation with immune cell infiltration levels. Additionally, single-cell RNA sequencing data from the GEO 148071 further confirmed this hypothesis.

Our study first established a clinical model of vitamin metabolism and evaluated its prognostic value in LUAD. However, there are certain limitations that must be acknowledged. Firstly, our study was retrospective and relied solely on data obtained from three public datasets. Besides, it is difficult to conduct larger-scale studies in the real-world to validate the findings. Secondly, because our prognostic model is a multi-parameter and multi-factor model, it is difficult to verify through specific experiments.

Thus, it is critical to undertake further research to verify the model in clinical practice.

Conclusions

We have identified and demonstrated the significance of VRS in LUAD for the first time, which can serve as reliable predictors of survival and guide personalized treatment strategies for LUAD patients.

Acknowledgments

Funding: This study was supported by the Fund of Nature Science Foundation of Hunan Province (No. 2023JJ40828); National Natural Science Foundation, China (No. 82150006); Chinese Society of Clinical Oncology Research Foundation (Nos. Y-HR2017-117, Y-HH202102-0060); Beijing Medical and Health Foundation (No. YWJKJJHKYJJ-F3046D); Qujiang District Quzhou City Life Oasis Public Service Center (No. BJHA-CRP-040); Wu Jieping Medical Foundation (No. 320.6750.2023-05-10).

Footnote

Reporting Checklist: The authors have completed the TRIPOD reporting checklist. Available at https://tlcr.amegroups.com/article/view/10.21037/tlcr-24-245/rc

Peer Review File: Available at https://tlcr.amegroups.com/article/view/10.21037/tlcr-24-245/prf

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at https://tlcr.amegroups.com/article/view/10.21037/tlcr-24-245/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).

References

1. Siegel RL, Miller KD, Wagle NS, et al. Cancer statistics, 2023. CA Cancer J Clin 2023;73:17-48.
2. Travis WD, Brambilla E, Nicholson AG, et al. The 2015 World Health Organization Classification of Lung Tumors: Impact of Genetic, Clinical and Radiologic Advances Since the 2004 Classification. J Thorac Oncol 2015;10:1243-60.
3. Allemani C, Matsuda T, Di Carlo V, et al. Global surveillance of trends in cancer survival 2000-14 (CONCORD-3): analysis of individual records for 37 513 025 patients diagnosed with one of 18 cancers from 322 population-based registries in 71 countries. Lancet 2018;391:1023-75.
4. Li MY, Liu LZ, Dong M. Progress on pivotal role and application of exosome in lung cancer carcinogenesis, diagnosis, therapy and prognosis. Mol Cancer 2021;20:22.
5. Duma N, Santana-Davila R, Molina JR. Non-Small Cell Lung Cancer: Epidemiology, Screening, Diagnosis, and Treatment. Mayo Clin Proc 2019;94:1623-40.
6. Nur SM, Rath S, Ahmad V, et al. Nutritive vitamins as epidrugs. Crit Rev Food Sci Nutr 2021;61:1-13.
7. Vincenzetti S, Santini G, Polzonetti V, et al. Vitamins in Human and Donkey Milk: Functional and Nutritional Role. Nutrients 2021;13:1509.
8. Liu N, Li X, Fu Y, et al. Inhibition of lung cancer by vitamin D depends on downregulation of histidine-rich calcium-binding protein. J Adv Res 2021;29:13-22.
9. Wang W, Hu W, Xue S, et al. Vitamin D and Lung Cancer; Association, Prevention, and Treatment. Nutr Cancer 2021;73:2188-200.
10. Rozmus D, Ciesielska A, Płomiński J, et al. Vitamin D Binding Protein (VDBP) and Its Gene Polymorphisms-The Risk of Malignant Tumors and Other Diseases. Int J Mol Sci 2020;21:7822.
11. McFarland DC, Fernbach M, Breitbart WS, et al. Prognosis in metastatic lung cancer: vitamin D deficiency and depression-a cross-sectional analysis. BMJ Support Palliat Care 2022;12:339-46.
12. Min YN, Niu ZY, Sun TT, et al. Vitamin E and vitamin C supplementation improves antioxidant status and immune function in oxidative-stressed breeder roosters

by up-regulating expression of GSH-Px gene. Poult Sci 2018;97:1238-44.
13. Bivona JJ 3rd, Patel S, Vajdy M. Induction of cellular and molecular Immunomodulatory pathways by vitamin E and vitamin C. Expert Opin Biol Ther 2017;17:1539-51.
14. Panebianco C, Eddine FBN, Forlani G, et al. Probiotic Bifidobacterium lactis, anti-oxidant vitamin E/C and anti-inflammatory dha attenuate lung inflammation due to pm2.5 exposure in mice. Benef Microbes 2019;10:69-75.
15. Maggini S, Wintergerst ES, Beveridge S, et al. Selected vitamins and trace elements support immune function by strengthening epithelial barriers and cellular and humoral immune responses. Br J Nutr 2007;98 Suppl 1:S29-35.
16. de Rouw N, Piet B, Derijks HJ, et al. Mechanisms, Management and Prevention of Pemetrexed-Related Toxicity. Drug Saf 2021;44:1271-81.
17. Barrett T, Wilhite SE, Ledoux P, et al. NCBI GEO: archive for functional genomics data sets--update. Nucleic Acids Res 2013;41:D991-D995.
18. Hutter C, Zenklusen JC. The Cancer Genome Atlas: Creating Lasting Value beyond Its Data. Cell 2018;173:283-5.
19. 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.
20. Schabath MB, Welsh EA, Fulp WJ, et al. Differential association of STK11 and TP53 with KRAS mutation-associated gene expression, proliferation and immune surveillance in lung adenocarcinoma. Oncogene 2016;35:3209-16.
21. Ouyang W, Jiang Y, Bu S, et al. A Prognostic Risk Score Based on Hypoxia-, Immunity-, and Epithelialto-Mesenchymal Transition-Related Genes for the Prognosis and Immunotherapy Response of Lung Adenocarcinoma. Front Cell Dev Biol 2022;9:758777.
22. Jiang Y, Ouyang W, Zhang C, et al. Prognosis and Immunotherapy Response With a Novel Golgi Apparatus Signature-Based Formula in Lung Adenocarcinoma. Front Cell Dev Biol 2021;9:817085.
23. Dib L, San-Jose LM, Ducrest AL, et al. Selection on the Major Color Gene Melanocortin-1-Receptor Shaped the Evolution of the Melanocortin System Genes. Int J Mol Sci 2017;18:2618.
24. Han Y, Wang Y, Dong X, et al. TISCH2: expanded datasets and new tools for single-cell transcriptome analyses of the tumor microenvironment. Nucleic Acids Res 2023;51:D1425-31.
25. Yang W, Soares J, Greninger P, et al. Genomics of Drug Sensitivity in Cancer (GDSC): a resource for therapeutic biomarker discovery in cancer cells. Nucleic Acids Res 2013;41:D955-61.
26. Oliver AL. Lung Cancer: Epidemiology and Screening. Surg Clin North Am 2022;102:335-44.
27. Peterson CT, Rodionov DA, Osterman AL, et al. B Vitamins and Their Role in Immune Regulation and Cancer. Nutrients 2020;12:3380.
28. Magrì A, Germano G, Lorenzato A, et al. High-dose vitamin C enhances cancer immunotherapy. Sci Transl Med 2020;12:eaay8707.
29. Negri M, Gentile A, de Angelis C, et al. Vitamin D-Induced Molecular Mechanisms to Potentiate Cancer Therapy and to Reverse Drug-Resistance in Cancer Cells. Nutrients 2020;12:1798.
30. Carlberg C, Velleuer E. Vitamin D and the risk for cancer: A molecular analysis. Biochem Pharmacol 2022;196:114735.
31. Reda M, Ngamcherdtrakul W, Gu S, et al. PLK1 and EGFR targeted nanoparticle as a radiation sensitizer for non-small cell lung cancer. Cancer Lett 2019;467:9-18.
32. Archambault V, Glover DM. Polo-like kinases: conservation and divergence in their functions and regulation. Nat Rev Mol Cell Biol 2009;10:265-75.
33. Barr FA, Silljé HH, Nigg EA. Polo-like kinases and the orchestration of cell division. Nat Rev Mol Cell Biol 2004;5:429-40.
34. Hannon GJ, Casso D, Beach D. KAP: a dual specificity phosphatase that interacts with cyclin-dependent kinases. Proc Natl Acad Sci U S A 1994;91:1731-5.
35. Nalepa G, Barnholtz-Sloan J, Enzor R, et al. The tumor suppressor CDKN3 controls mitosis. J Cell Biol 2013;201:997-1012.
36. Lai MW, Chen TC, Pang ST, et al. Overexpression of cyclin-dependent kinase-associated protein phosphatase enhances cell proliferation in renal cancer cells. Urol Oncol 2012;30:871-8.
37. Yu C, Cao H, He X, et al. Cyclin-dependent kinase inhibitor 3 (CDKN3) plays a critical role in prostate cancer via regulating cell cycle and DNA replication signaling. Biomed Pharmacother 2017;96:1109-18.
38. Lee SW, Reimer CL, Fang L, et al. Overexpression of kinase-associated phosphatase (KAP) in breast and prostate cancer and inhibition of the transformed phenotype by antisense KAP expression. Mol Cell Biol 2000;20:1723-32.
39. Zang X, Chen M, Zhou Y, et al. Identifying CDKN3

Gene Expression as a Prognostic Biomarker in Lung Adenocarcinoma via Meta-analysis. Cancer Inform 2015;14:183-91.
40. Letson C, Padron E. Non-canonical transcriptional consequences of BET inhibition in cancer. Pharmacol Res 2019;150:104508.
41. Shen S, Li K, Liu Y, et al. Silencing lncRNA AGAP2-AS1 Upregulates miR-195-5p to Repress Migration and Invasion of EC Cells via the Decrease of FOSL1 Expression. Mol Ther Nucleic Acids 2020;20:331-44.
42. Han YH, Mun JG, Jeon HD, et al. Betulin Inhibits Lung Metastasis by Inducing Cell Cycle Arrest, Autophagy, and Apoptosis of Metastatic Colorectal Cancer Cells. Nutrients 2019;12:66.
43. Asghar U, Witkiewicz AK, Turner NC, et al. The history and future of targeting cyclin-dependent kinases in cancer therapy. Nat Rev Drug Discov 2015;14:130-46.
44. He J, Huang Z, Han L, et al. Mechanisms and management of 3rd-generation EGFR-TKI resistance in advanced non-small cell lung cancer (Review). Int J Oncol 2021;59:90.
45. Weiss JM, Csoszi T, Maglakelidze M, et al. Myelopreservation with the CDK4/6 inhibitor trilaciclib in patients with small-cell lung cancer receiving first-line chemotherapy: a phase Ib/randomized phase II trial. Ann Oncol 2019;30:1613-21.
46. Gao X, Leone GW, Wang H. Cyclin D-CDK4/6 functions in cancer. Adv Cancer Res 2020;148:147-69.
47. Iksen, Pothongsrisit S, Pongrakhananon V. Targeting the PI3K/AKT/mTOR Signaling Pathway in Lung Cancer: An Update Regarding Potential Drugs and Natural Products. Molecules 2021;26:4100.
48. Zhang Y, Lv P, Ma J, et al. Antrodia cinnamomea exerts an anti-hepatoma effect by targeting PI3K/AKT-mediated cell cycle progression in vitro and in vivo. Acta Pharm Sin B 2022;12:890-906.
49. Matsushime H, Ewen ME, Strom DK, et al. Identification and properties of an atypical catalytic subunit (p34PSK-J3/cdk4) for mammalian D type G1 cyclins. Cell 1992;71:323-34.
50. Yang L, Zhou Y, Li Y, et al. Mutations of p53 and KRAS activate NF-κB to promote chemoresistance and tumorigenesis via dysregulation of cell cycle and suppression of apoptosis in lung cancer cells. Cancer Lett 2015;357:520-6.
51. Kuncman Ł, Orzechowska M, Milecki T, et al. High FLT3 expression increases immune-cell infiltration in the tumor microenvironment and correlates with prolonged disease-free survival in patients with non-small cell lung cancer. Mol Oncol 2024;18:1316-26.

Cite this article as: Chen Y, Jiang Y, Li X, Huang H, Zhou Y, Zhang C, Wang S, Bohnenberger H, Gao Y. Identification of a novel prognostic signature based on vitamin metabolism clustering-related genes in lung adenocarcinoma. Transl Lung Cancer Res 2024;13(5):1084-1100. doi: 10.21037/tlcr-24-245

Supplementary

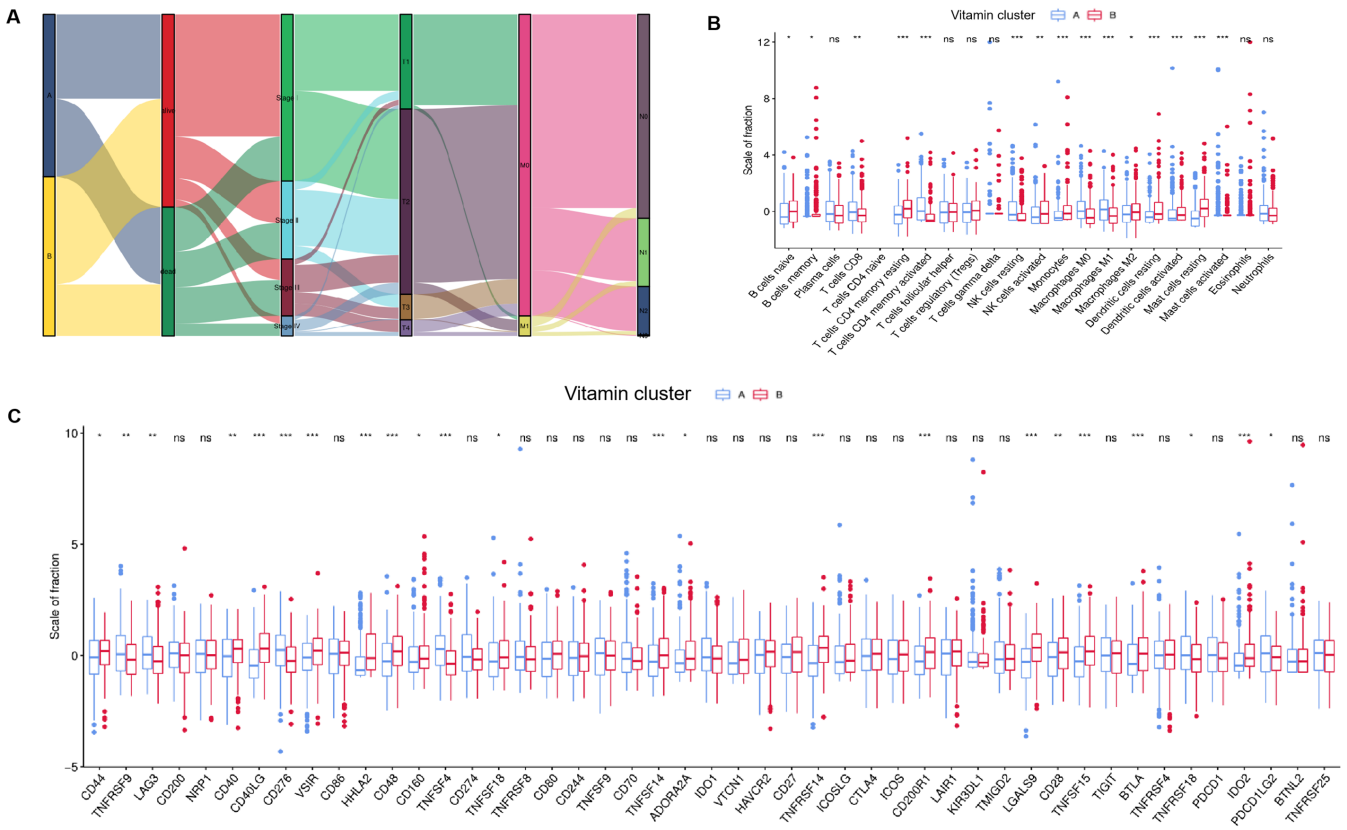

Figure S1 Analysis of clinical characteristics (A), immune cell infiltration (B) and immune checkpoint (C) between the 2 unsupervised clustering groups. *, P<0.05; **, P<0.01; ***, P<0.001. NK, natural killer; ns, not significant.

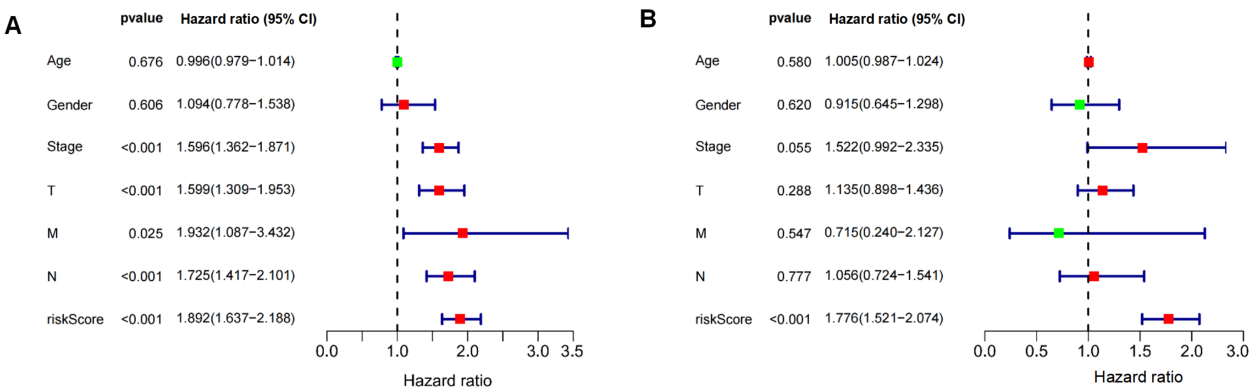

Figure S2 Univariate Cox regression (A) and multivariate Cox regression (B) analysis of VRS and clinical factors. CI, confidence interval; VRS, vitamin metabolism gene-related risk score.

https://dx.doi.org/10.21037/tlcr-24-245