## Identification of metabolism-associated molecular classification for effect and prognosis in lung adenocarcinoma based on multidatabases including the cancer genome atlas and gene expression omnibus

SAGE Open Medicine Volume 13: 1–15 © The Author(s) 2025 Article reuse guidelines: sagepub.com/journals-permissions DOI: 10.1177/20503121251341114 journals.sagepub.com/home/smo



## Lilin Que<sup>1\*</sup>, Zhibing Liu<sup>2\*</sup>, Yinghui Wu<sup>3</sup>, Lan Luo<sup>1</sup> and Leifeng Liang<sup>1</sup>

#### Abstract

**Background:** Lung adenocarcinoma is a highly heterogeneous group of diseases with distinct molecular genetic features, pathological characteristics, metabolic profiles, and clinical behaviors. However, the clinical relevance of metabolic characteristics of lung adenocarcinoma remains unclear. This study aimed to describe the molecular characteristics of lung adenocarcinoma. **Methods:** The gene expression profiles of 1037 lung adenocarcinoma samples were downloaded from The Cancer Genome Atlas and Gene Expression Omnibus databases. This study is based on sample data from 2006 to 2020. The long-time span and sufficient sample size ensure the robustness of the research findings. Using unsupervised transcriptome analysis, we identified three distinct subtypes (C1, C2, and C3). We then compared the prognostic traits, transcriptome characteristics, metabolic signatures, immune infiltration, clinical features, and drug sensitivity of the lung adenocarcinoma subclasses. A classifier was generated to determine lung adenocarcinoma classification, and we verified the clinical value of this classifier in other tumors. **Results:** Our results indicated that C1 possessed the most abundant metabolic pathways. Compared with C2 and C3, C1 possessed 35 metabolic pathways that exhibited significant differences. The immune score, matrix score, and immune infiltration for subtype C1 were significantly lower than those for subtypes C2 and C3, suggesting that C1 is a metabolically active subtype. Five metabolic pathways were observed in C2. Subtype C2 was associated with the best prognosis and exhibited the lowest tumor mutation burden and copy number variation. Subtype C3 comprised five metabolic pathways. Immune checkpoint analysis revealed that C3 cells may potentially benefit from immunotherapy.

**Conclusions:** Our study deepens the understanding of the metabolic characteristics of lung adenocarcinoma and may provide valuable information for immunotherapy.

#### **Keywords**

Lung adenocarcinoma, classification, metabolism, immune signatures

Received: 4 January 2025; accepted: 24 April 2025

## Introduction

Lung cancer continues to be the most dominant cause of cancer mortality worldwide, with more than 1 million deaths each year, and lung adenocarcinoma (LUAD) is the most common histological subtype of this disease.<sup>1–3</sup> LUAD is a highly heterogeneous group of diseases with distinct molecular genetic features, pathological characteristics, and clinical behaviors.<sup>4</sup> Despite advances in early diagnosis and new therapeutic strategies such as small molecule targeted therapy and immunotherapy that have provided new hope for patients with LUAD, the prognosis of LUAD patients remains far from satisfactory. Thus, there is

<sup>1</sup>Department of Oncology, The Sixth Affiliated Hospital of Guangxi Medical University, The First People's Hospital of Yulin, China <sup>2</sup>Department of Oncology, Binzhou Medical University Hospital, Shandong, China

<sup>3</sup>Department of Pathology, The Sixth Affiliated Hospital of Guangxi Medical University, The First People's Hospital of Yulin, China

\*These authors have contributed equally to this work and share first authorship.

#### **Corresponding author:**

Leifeng Liang, Department of Oncology, The Sixth Affiliated Hospital of Guangxi Medical University, The First People's Hospital of Yulin, Jiaoyu Road, Yuzhou District, Yulin, Guangxi 537000, China. Email: liangleifeng@stu.gxmu.edu.cn

Creative Commons Non Commercial CC BY-NC: This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 License (https://creativecommons.org/licenses/by-nc/4.0/) which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (https://us.sagepub.com/en-us/nam/open-access-at-sage).

Characteristics	TCGA-LUAD (n=487)	GSE30219 (n=85)	GSE31210 (n=226)	GSE37745 (n = 106)	GSE42127 (n = 133)
Age	66 (59–72)	60 (55–69)	61 (55–65)	64 (55–70)	66 (59–74)
Pack years smoked	40 (21.5–51)	NA	NA	NA	NA
Year smoked	32.5 (20.5–40)	NA	NA	NA	NA
Overall survival	, , , , , , , , , , , , , , , , , , ,				
Alive	310	40	191	29	90
Dead	177	45	35	77	43
Gender					
Female	265	19	121	60	65
Male	222	66	105	46	68
Pathological stage					
I	260	NA	168	70	89
II	116	NA	58	19	22
111	79	NA	0	13	20
IV	25	NA	0	4	I
Missing	7	NA	0	0	I

Table I. Clinical characteristics of patients with LUAD in TCGA and GEO sets.

a need to identify the molecular mechanisms that contribute to LUAD to develop new and effective prevention and treatment strategies.

With the development of RNA-sequencing technology and microarrays, gene expression profiling has emerged as a useful tool for classifying tumors.<sup>5–8</sup> For example, Hu et al.<sup>9</sup> used *k*-means clustering, self-organizing map neural networking, and hierarchical clustering methods to define four distinct subtypes of LUAD. Different subtypes yield distinct prognoses. Patients with subtype four exhibited one of the lowest mortality rates. Subtypes 1 and 2 were highly enriched in immunity-related biological processes. These results confirm that dysregulation of the immune system plays an important role in the pathogenesis of LUAD.

Dysregulated metabolism is indispensable for cancer cell proliferation.<sup>10-12</sup> Abnormal cancer metabolism leads to unique metabolic dependencies that can be targeted for therapeutic effects.<sup>13,14</sup> Based on this, we believe that insights into the differences in metabolism of various LUAD subtypes may lead to the discovery of new treatment modalities. Recently, a study divided colorectal cancer samples into three subclasses according to metabolic genes: metabolic active subtype (C1), metabolic exhausted subtype (C2), and intermediate metabolic activity subtype (C3).<sup>15</sup> The molecular, immune, and clinical characteristics of each subtype are different. However, molecular classification of LUAD metabolism has not yet been reported. In this study, we classified LUAD from a metabolic perspective. We compared the prognosis characteristics, transcriptome characteristics, metabolic signatures, immune infiltration, clinical features, and drug sensitivities of the LUAD subclasses. A classifier was generated to determine LUAD classification, and we verified the clinical value of this classifier in other tumors.

### Materials and methods

#### Data preprocessing

Clinical and molecular data of LUAD were collected from The Cancer Genome Atlas (TCGA)<sup>16</sup> (https://cancergenome. nih.gov/) and Gene Expression Omnibus (GEO) databases<sup>17</sup> (https://www.ncbi.nlm.nih.gov/geoprofiles/), and only tumor samples were retained. The TCGA-LUAD datasets were downloaded using the TCGABiolinks package<sup>18</sup> as Fragments Per Kilobase of transcript per Million mapped reads (FPKM) values. FPKM was then transformed into Transcripts Per Million according to the GENCODE version 27 annotation file. After data processing, 487 patients with LUAD from the TCGA-LUAD project were included in the training study. To verify this, expression data of human LUAD mRNA were downloaded from the GEO database (https://www.ncbi. nlm.nih.gov/geo/). Four datasets, GSE3021919 (containing 85 LUAD samples), GSE31210<sup>20,21</sup> (226 LUAD samples), GSE37745<sup>22–24</sup> (106 LUAD samples), and GSE42127<sup>25,26</sup> (133 LUAD samples), were selected as the testing sets. Clinical information was extracted from the TCGA pan-cancer clinical data resource. The clinical characteristics of the 1037 patients are presented in Table 1. Gene copy number data were obtained using Firehose.

## Identification of LUAD subtypes

Nonnegative matrix factorization (NMF) is an effective matrix decomposition method that decomposes a large nonnegative matrix into several small matrices to achieve clustering and typing.<sup>27</sup> In our study, we prepared 2752 metabolism-relevant genes for subsequent NMF clustering. To reasonably classify LUAD samples, we first used the ComBat algorithm to eliminate batch effects across different cohorts. Candidate genes with an expression value of zero

and a low median absolute deviation value  $\leq 0.5$  across the patients with LUAD were excluded from the study. A Cox proportional hazards model was also created using the "survival" R package to screen for meaningful genes for overall survival (p < 0.1). Subsequently, unsupervised NMF was performed using the "NMF" R package on the training and testing datasets to identify LUAD subtypes with distinct metabolic profiles.<sup>28</sup> We assessed clustering stability using the cophenetic correlation coefficient. The optimal number of clusters was selected as K, beginning to fall within the magnitude of the cophenetic correlation coefficient.<sup>29</sup>

To evaluate the stability of LUAD subtypes across different datasets and explore their relationship with existing molecular classifications, we utilized subclass Mapping (SubMap) analysis (Gene Pattern). We established a reference model using the gene expression data of LUAD subtypes from the training set and matched the expression matrix of the test dataset to this reference model to assess the consistency between different subtypes. Furthermore, we applied SubMap analysis to predict the potential response of different subtypes to immune checkpoint inhibitor therapy (e.g., PD-1 inhibitors) to explore the association between LUAD subtypes and immunotherapy sensitivity.

Additionally, we used t-distributed stochastic neighbor embedding (t-SNE) analysis to validate subtype assignments based on the mRNA expression data, ensuring the scientific rigor of the classification.

#### Gene set variation analysis

Gene set variation analysis (GSVA) is a nonparametric and unsupervised gene set enrichment method that can calculate a certain pathway or signature scores based on transcriptomic data.<sup>30</sup> Data of the 115 metabolism-associated gene signatures and 11 cancer pathways were obtained from previously published studies.<sup>31,32</sup> The GSVA R package was used to identify differences among different gene sets. Subsequently, differential expression analysis was conducted using linear models for microarray analysis (LIMMA) software (R, Bioconductor),<sup>33</sup> and differentially expressed signatures were defined as those with an absolute log2 fold change (FC) > 0.2 (adjusted p < 0.05).

#### Estimation of immune infiltration

The absolute abundance of eight immune and two nonimmune stromal cell populations was estimated using microenvironment cell populations-counter (MCP-counter), an independent bioinformatics tool to assess immune cell enrichment.<sup>34</sup> Furthermore, the single-sample GSEA algorithm was another approach used to estimate immune infiltration in this study.<sup>35</sup> An additional six immune cell populations were estimated using the Bioconductor R package GSVA. Additionally, the ESTIMATE<sup>36</sup> algorithm was used to calculate immune and stromal scores, which can represent enrichment of stromal and immune cell gene signatures.

## Characterization of LUAD subtypes

The LIMMA package was used for differentially expressed gene (DEG) analysis according to normalized count data. DEGs were defined as absolute  $\log 2 \text{ FC} > 1$  (corrected p < 0.01). Gene set files for "c2.cp.KEGG.v6.2. symbols" downloaded "h.all.v6.2. symbols" and from the Molecular Signatures Database were used to perform pathway and functional enrichment analyses using the CLUSTERPROFILER R package,<sup>37</sup> and the significance threshold was set at adjusted p < 0.05. We used nearest template prediction (NTP) analyses (Gene Pattern modules) to predict previously published LUAD molecular classifications and then compared the prediction results with our classification.

## Generation of the classifier and performance validation

We defined statistically significant differential genes as adjusted p < 0.01 and absolute log2 FC>2. Only genes with significant differences in all three possible comparisons were considered as subtype-specific genes. We selected the top 30 genes with the largest log2 FC values in each cohort to develop a prediction model and generate our 90-gene classifier. We then repeated the subclass prediction utilizing the 90-gene signature on the validation sets using the NTP algorithm.

# Prediction of the benefit of each subclass from immunotherapy

Based on the 90-gene classifier, a consistency check was performed using the NTP algorithm to predict the metabolism-related classes for each sample. The data from patients with LUAD who received immunotherapy were used to indirectly predict the immunotherapy efficacy of our subtypes by measuring the similarity of gene expression profiles between our subtypes and patients with LUAD based on SubMap analyses.

### Prediction of drug sensitivity in each subclass

To identify effective antitumor drugs, we downloaded drug sensitivity data from the Genomics of Drug Sensitivity in the Cancer (GDSC) database. The Kruskal–Wallis test was used to compare the sensitivity of 100 drugs in the GDSC database. IC<sub>50</sub> data were obtained from the GDSC database. According to the ranking of cell lines from low to high IC<sub>50</sub>, the top 1/3rd of the cell lines were defined as drug-sensitive, and the last 1/3rd of the cell lines were defined as drug-resistant.

## GO and KEGG enrichment analysis

ClusterProfiler software was used for enrichment analyses that included GO and KEGG enrichment analyses of DEGs



**Figure 1.** Identification of lung adenocarcinoma (LUAD) subclasses using nonnegative matrix factorization (NMF) consensus clustering in the metadata set. (a) Principal component analysis plot of the combined expression profile of cohort data. (b) NMF clustering using 816 metabolism-associated genes. (c) t-distributed stochastic neighbor embedding (t-SNE) analysis supported the stratification into three subclasses. (d–h) OS of three subclasses in training set and testing sets.

among the three subtypes (adj. p < 0.05, FDR  $\leq 0.05$ ). Subsequent bar charts and bubble plots were generated using the R package ggplot2 and used for statistical analysis and data visualization.

### Pan-cancer analysis of 90-gene classifier

The gene expression profiling interactive analysis (GEPIA) database was used to study the 90-gene classifier. A heatmap was used to display the expression levels in the tumor tissues of individual cancers in TCGA.<sup>38</sup>

## Prediction of transcription factors

Considering that transcription factors exert various functions in the context of gene regulation, we analyzed transcription factors acquired from the transcriptional regulatory relationships unraveled by sentence-based text-mining (TRRUST) database, which are most likely to regulate the 90 genes included in the classifier.<sup>39</sup>

#### Statistical analysis

All data processing and analyses were performed using R (version 4.0.2) and Excel (Microsoft). In this study, we used several statistical methods to analyze the differences among LUAD subtypes (C1, C2, and C3). Prior to intergroup difference analysis, we first calculated the variance for each group

to assess the degree of data dispersion. Differences between two groups were analyzed using Student's *t*-test for normally distributed data or the Mann–Whitney *U*-test for nonnormally distributed data. Kruskal–Wallis test (for nonparametric data) was employed for multiple group comparisons. Contingency table data were compared using the Chi-squared test or Fisher's exact test. Survival analysis was performed using the Kaplan–Meier method, and the log-rank test was used to assess the prognostic differences among different metabolic subtypes. The hazard ratio was estimated using a univariate Cox proportional hazard regression model. We considered two-tailed p < 0.05 as statistically significant.

## Results

### NMF identifies three subtypes in LUAD

The clinical characteristics of patients from different cohorts are listed in Table 1. After removing batch effects, we created a principal component analysis (PCA) plot (Figure 1(a)). In the results of PCA, the first principal component (PC1) accounted for 8.8% of the variance, indicating a substantial contribution of this component to the data distribution. Based on 2752 previously reported metabolically related genes, univariable Cox regression was used to filter genes related to Overal Sutvival (OS) time (p < 0.1), and 517 candidate genes were screened out. The dataset for LUAD samples from TCGA was clustered according to the 517 candidate

genes using NMF consensus clustering. The response profiles were clustered using NMF (Figure 1(b)). To evaluate and choose from the optimal clustering scheme, cophenetic correlation coefficients were calculated, and k=3 was chosen as the optimal number of clusters after a comprehensive consideration. We identified three clusters using NMF consensus clustering that included C1 (n=113), C2 (n=103), and C3 (n=234). t-SNE was used to validate the initial cluster analysis (Figure 1(c)). The two-dimensional t-SNE distribution patterns were robustly consistent with our subtype clustering. Next, unsupervised clustering analysis was performed after merging the remaining four datasets (GSE30219, GSE31210, GSE37745, and GSE42127). This analysis resulted in three subtypes: C1 (n=93), C2 (n=98), and C3 (n=98). We then performed survival analyses across the five LUAD cohorts. The results revealed that the OS of the C2 subtype was significantly longer than those of the C1 and C3 subtypes in the TCGA-LUAD cohort (p=0.002). A similar prognosis was observed in the validation dataset (Figure 1(d)–(h)).

## Correlation between the LUAD subtypes and metabolism-related signatures

Considering that the classification was based on metabolismrelated genes, we further explored the unique metabolic features of each subtype. First, we used the R package GSVA to calculate the expression levels of genes associated with metabolism and carcinogenesis. To define the subtype-specific differential metabolic pathways, we used |logFC| = 0.2and padj=0.05 and plotted a heat map (Figure 2(a)). The results revealed that most of the differential metabolic pathways were enriched in C1. Compared to C2 and C3, C1 possessed 35 significantly different metabolic pathways, including amino acids, lipids, and other metabolism-related signatures. This clearly indicated that C1 was the most metabolically active subtype. Additionally, C2 has five enriched metabolic pathways, primarily related to lipid metabolism, while C3 also has five metabolic pathways, but with lower enrichment levels compared to C1.

To further understand the characteristics of these subtypes, 11 carcinogenesis pathways were evaluated and quantified using the GSVA algorithm (Figure 2(b)). The results revealed that the NORCH signature of C1 was significantly higher than that of C2 and C3. Additionally, C3 displayed higher expression of PI3K and cell cycle pathways, while C2 was significantly enriched in HIPPO, TGF- $\beta$ , RTK/RAS, TP53, WNT, and angiogenesis pathways. The results presented above indicate that this classification may be strongly associated with cancer.

To evaluate the heterogeneity among the three subtypes, we used the ESTIMATE algorithm to calculate the stromal and immune scores and construct violin and box plots (Figure 2(c)and (d)). The results indicated that the immune scores of C2 and C3 were significantly higher than that of C1 (p < 0.001, Figure 2(c)), suggesting that C1 may have lower

immune infiltration; however, there was no significant difference between C2 and C3 in terms of immune scores. Similar to the immune score, the stromal scores analysis also revealed that the stromal content in C1 was significantly lower than that in C2 and C3 (p < 0.00001), further supporting the notion that C1 represents a subtype with high metabolic activity, but a weaker immune microenvironment, whereas C2 and C3 may play a more significant role in immune response (p < 0.00001, Figure 2(d)). Overall, distinct LUAD subtypes exhibit significant differences in metabolic profiles, cancerrelated signaling pathways, and immune microenvironments, which may influence their clinical manifestations and therapeutic responses.

## The correlation between LUAD subtypes and immune infiltration

The results presented above indicate that there was a significant difference in immune scores among subtypes. To further investigate the immune microenvironment characteristics of LUAD subtypes, we analyzed the infiltration of immune cells and the expression levels of immune checkpoint genes to describe the immunologic landscape in the TCGA-LUAD database. We estimated the abundance of 16 immune cell types using the MCP-counter1 and single-sample GSEA algorithms and presented them in a heat map of immune functionrelated genes (Figure 3(a)). We then mapped a box plot to reflect the differences among the three groups. The overall levels of immune cell infiltration were significantly higher in C2 and C3 compared to C1. Specifically, T cells (CD8<sup>+</sup> T cells), B cells, and activated NK cells were more highly expressed in C2 and C3. In contrast, C1 exhibited lower levels of immune cell infiltration, suggesting that C1 may possess an immune-cold characteristic. Figure 3(b) further validated the enrichment of different immune cell subsets across subtypes. CD8<sup>+</sup> T cells and NK cells were most highly enriched in C3 (p < 0.001), potentially enhancing antitumor immune activity. Additionally, C2 was enriched for B cells and memory T cells, which may enhance antigen-presenting functions. In contrast, C1 exhibited higher levels of M2 macrophages (p < 0.05), indicating a stronger immunosuppressive microenvironment consistent with its low immune score.

To further validate the immune microenvironment characteristics, we analyzed the expression of 13 immune checkpoint genes across different subtypes (Figure 3(c)). Those genes were selected based on drug inhibitors currently approved for specific cancer types. The analysis revealed that the expression levels of PD-1 (PDCD1), PD-L1 (CD274), CTLA4, and LAG3 were significantly higher in C3 compared to C1 and C2 (p < 0.001), suggesting that this subtype may be more responsive to immune checkpoint inhibitors. In contrast, C1 exhibited the lowest expression of immune checkpoint genes, further confirming its "immunecold" status. C2 had intermediate expression levels of CD86 and ICOS between C1 and C3, which may imply that it could



**Figure 2.** Association between lung adenocarcinoma (LUAD) subtypes and metabolism-related signatures. (a) Heatmap of GSVA enrichment scores for metabolic pathways, showing the expression levels of metabolic gene sets in the C1, C2, and C3 subtypes. (b) Boxplot of GSVA results for 11 cancer-related signaling pathways. Boxplot of immune scores (c) and stromal scores (d) from ESTIMATE of subtypes. *p*-values are labeled above each boxplot with asterisks. ns: no significance.

\*p < 0.05,\*\*p < 0.01, \*\*\*p <0.001

participate in antitumor immune responses under certain conditions, but its direct response to immunotherapy may not be as pronounced as that of C3.

In summary, the analysis of immune cell infiltration and immune checkpoint gene expression uncovered distinct immune profiles among different LUAD subtypes. C3 exhibited an "immune-hot" tumor phenotype, characterized by high levels of immune cell infiltration and immune checkpoint gene expression, and may represent a subgroup that could potentially benefit from immunotherapy. C1 displayed an "immune-cold" phenotype, which may require combination therapeutic strategies to enhance sensitivity to immunotherapy, while C2 had an intermediate immune profile between C1 and C3.

## Correlation between LUAD subgroups and clinical characteristics in TCGA and GEO datasets

To investigate the distribution of LUAD subtypes across various clinical characteristics, we analyzed patient information from the TCGA-LUAD dataset as well as the GSE30219, GSE31210, GSE37745, and GSE42127 datasets. Figure 4(a) presents the stratified statistics of the TCGA-LUAD cohort, including patient age, smoking status, pathological stage (pStage), mTOR pathway activation, and DNA methylation profiles. The overall trend indicates that C1 is overrepresented in patients with high smoking exposure (>30 pack-years), mTOR pathway activation, and specific methylation patterns, suggesting that C1 may be characterized



**Figure 3.** Immune characteristics of three subtypes in The Cancer Genome Atlas (TCGA)-lung adenocarcinoma (LUAD) set. (a) Heatmap describing the abundance of immune and stromal cell populations in C1, C2, and C3. Immune cell scores were calculated based on MCP-counter and single-sample GSEA (ssGSEA). (b) Boxplot of the abundance of immune and stromal cell populations distinguished by different subtypes. (c) Expression levels of 13 immune checkpoint genes. Statistical comparisons were performed using the Kruskal–Wallis test, with significance levels indicated in the figures. \*p < 0.05, \*\*p < 0.01, \*\*\*p < 0.001.

by unique metabolic and signaling network reprogramming. In contrast, C2 is more prevalent among patients with lower smoking exposure and is associated with lower levels of mTOR pathway activation and distinct DNA methylation profiles. C3 exhibits intermediate distribution across these clinical features.

To further validate whether the clinical characteristics of different LUAD subtypes are consistent across independent GEO datasets, we repeated the classification analysis in the GSE30219, GSE31210, GSE37745, and GSE42127 datasets (Figure 4(b)–(e)). The results from the GEO cohorts were largely consistent with those from the TCGA-LUAD dataset, indicating that C1 is more common in patients with advanced pathological stages (Stages III and IV) and heavy smoking history, while C2 is relatively more prevalent in patients with early-stage disease (Stages I and II). Additionally, the enrichment of the mTOR signaling pathway showed a similar trend

across the GEO datasets, with C1 exhibiting higher levels of mTOR pathway activation and C2 showing the lowest levels of mTOR activation.

These findings demonstrate good cross-dataset consistency in the distribution of LUAD subtypes across clinical and molecular features, suggesting that C1 may be driven by metabolic factors, while C2 may be associated with more primary features of LUAD. This subtype-specific molecular pattern may hold significant implications for LUAD diagnosis, stratification, and the development of personalized therapeutic strategies.

## Associations of LUAD subtypes with mutations, neoantigens, and copy number aberrations

To investigate the genetic mutations and genomic characteristics of LUAD metabolic subtypes, we analyzed the gene mutations, tumor mutation burden (TMB), predicted neoantigen load, and copy number variation (CNV) patterns across different subtypes (C1, C2, and C3) in the TCGA-LUAD dataset.

Figure 5(a) presents the mutation spectra of different LUAD subtypes, focusing on key driver genes with the highest mutation frequencies, such as *TP53*, *KRAS*, *NF1*, *COL5A1*, *BRAF*, and *ARID2*. Notably, the mutation frequency of *TP53* was significantly higher in C3 than in C1 and C2, reinforcing the characteristic of genomic instability in this subtype. As an important tumor suppressor gene, *TP53* mutations not only affect cell proliferation but are also associated with tumor sensitivity to immunotherapy. In contrast, C2 exhibited the lowest *TP53* mutation frequency, which may correlate with its relatively favorable prognosis.

Figure 5(b) and (c) further examines the differences in TMB and neoantigen load. Statistical analyses revealed that C3 had significantly higher TMB and neoantigen load compared to C1 and C2 (p < 0.001), while C2 exhibited the lowest TMB. These findings suggest that C3, with high TMB and neoantigen load, may be more likely to benefit from immunotherapy, as higher mutation burdens can generate more tumor neoantigens, thereby enhancing tumor immunogenicity.

Figure 5(d) and (e) evaluates the CNVs across the three subtypes, including gene amplifications and deletions. C3 exhibited the highest level of CNVs (p < 0.001), with more frequent amplification and deletion events involving multiple known driver genes. For instance, KRAS and MYC were more frequently amplified in C3. In contrast, C2 had the fewest CNVs, corresponding to its lower genomic instability.

Overall, the mutation analysis indicates that C3 is characterized by high mutation rates, neoantigen load, and genomic instability, and may benefit the most from ICI therapy. C1 exhibited intermediate levels of genomic variation, while C2 had the lowest overall mutation burden and genomic changes, consistent with its better clinical prognosis. These findings may guide personalized treatment decisions for LUAD patients.



**Figure 4.** Clinical characteristics of lung adenocarcinoma (LUAD) subtypes in the TCGA cohort and GEO testing cohort. (a) Distribution of subtypes (C1, C2, and C3) in the TCGA-LUAD dataset across demographic features, smoking status, pathway activation (mTOR), DNA methylation patterns, and pathological staging. (b–e) Analysis of clinical features of LUAD subtypes in GEO datasets: GSE30219 (b), GSE31210 (c), GSE37745 (d), and GSE42127 (e).



**Figure 5.** Association between lung adenocarcinoma (LUAD) subtypes and mutations, neoantigens, and copy number aberrations. (a) Waterfall map of driven carcinogenic mutations clustered in C1, C2, and C3. (b) Analysis of tumor mutation burden (TMB). (c) Analysis of neoantigen load. (d–e) Analysis of copy number variation (CNV), including gene amplification levels (d), and gene deletion levels (e).



**Figure 6.** GISTIC score analysis of the lung adenocarcinoma (LUAD) subtypes in The Cancer Genome Atlas (TCGA) cohort. Cytoband map of copy number alterations across subtypes CI (a), C2 (b), and C3 (c). Red indicates regions of amplification, and blue indicates regions of deletion. (d) Box plots of predicted  $IC_{50}$  values based on the Genomics of Drug Sensitivity in the Cancer (GDSC) database.

To further explore the CNVs associated with the three metabolic subtypes of LUAD, we employed GISTIC2.0 to analyze the genomic CNV patterns across the three groups of samples. Figure 6(a)–(c) illustrates the genomic copy number alteration profiles of C1, C2, and C3. C3 exhibited the most pronounced CNV alterations, including multiple amplifications and deletions across various loci, indicative of greater genomic instability. C1 showed significant amplifications at loci 1q, 8q, and 19q, and frequent deletions at loci 9p and 18q. In contrast, C2 had the fewest CNVs, typically limited to localized amplifications/deletions of a few driver genes, consistent with its lower TMB and better clinical prognosis.

## Prediction of drug susceptibility

To identify potential antineoplastic drugs associated with the ICI group, we downloaded drug response data for more than 100 agents from the GDSC database. The  $IC_{50}$  values of the

selected compounds across the different LUAD subtypes were compared using the Kruskal–Wallis test. We then listed the top 12 drugs with the most significant differences according to the *p*-values (Figure 6(d)).

## The 90-gene classifier and its performance verification

To establish a classifier for clinical use, subtype-specific genes were selected to develop a prediction model. Differential expression analysis of "subtype n versus other subtypes" was performed using the Limma package. After the Movics package analysis,<sup>40</sup> the first 30 genes with the highest log2 FC values in each subtype were selected for the development of the subtype classifier, and a correlation heat map was created (Figure 7(a)). Based on this result, we obtained a 90-gene classifier. The subtypes of TCGA and the test sets were then predicted using the NTP method. Heat maps were created to represent the degree of matching between the true and predicted subtypes (Figure 7(b) and (c)). The results revealed good consistency between the two separate methods (NMF and NTP), indicating that the 90-gene signature can reproducibly determine LUAD classification.

The observation that different subtypes exhibit different patterns of immune cell infiltration and expression levels of immune checkpoint genes indicates that further research is needed to predict immunotherapy responses. Subclass mapping was used to compare the degree of similarity in expression profiles between the three subtypes and the dataset containing 47 patients with lung cancer who received immunotherapy. The results suggested that patients in group C3 were most likely to respond to immunotherapy (Figure 7(d)).

### Functional enrichment analysis of gene classifiers

GO and KEGG functional enrichment analyses of the DEGs were conducted using the cluster profile R package. Visualization was performed using the R package ggplot2. The significance of enrichment for the top ten DEGs is presented in Figure S1 and Table S1.

GO enrichment analysis of the DEGs implicated numerous biological processes and pathways, including metabolic processes, immune-related responses, protein binding, major histocompatibility class II receptor activity, and multiple metabolic enzyme activities. The KEGG enrichment results indicated that genes were primarily enriched for cell cycle, antigen processing and presentation, protein digestion and absorption, phagosome, and cancer-related signaling pathways (Figure S2 and Table S2).

# Differences in expression of the 90-genes classifier in pan-cancer

The expression levels of the 90-genes in tumor tissues of individual cancers in the TCGA database were validated using GEPIA. The results are presented as a heat map representing DEGs among different cancer types (Figure S3(a)).

(a)





Figure 7. Identification of predictive classifier and putative targeted therapeutic and immunotherapeutic responses. (a) Heatmap of the expression levels of the 90-gene classifier. Concordance of The Cancer Genome Atlas (TCGA)-lung adenocarcinoma (LUAD) (b) and testing cohort (c) prediction between the 90-gene classifier and the original prediction based on nonnegative matrix factorization. (d) The predicted molecular targeted therapy and immunotherapy response of the classifier.

## Prediction of transcription factors that can regulate the 90-genes

The transcription factors of the 90-genes were acquired from the TRRUST database. We identified 34 transcription factors that potentially regulate the 90-genes (Table.S3), including multiple tumor-associated genes, proto-oncogenes, and interferon regulatory factors. The transcription factors and their targets described above were analyzed using the Metascape platform and are indicated with an enrichment bar chart

(Figure S3(b) and (c)). A molecular interactive network was then constructed for the 90-genes, with colors representing the strength of significance (Figure S3(d)).

## Discussion

Dysregulation of cellular metabolism in cancer cells is indispensable for indefinite proliferation of cancer cells and represents a hallmark of cancer.<sup>10-12</sup> The rewiring of cellular metabolism results in a unique set of metabolic phenotypes that (1) allow for earlier cancer diagnosis, (2) better predict cancer risk, (3) guide therapy selection, and (4) facilitate the development of methods to monitor therapeutic effectiveness. Changes in the metabolism of cancer cells can lead to unique metabolic dependencies that provide an excellent opportunity for targeted therapy.<sup>13,14</sup> To identify the LUAD subsets associated with metabolic processes and good prognosis, we classified the metabolic spectrum of LUAD samples comprehensively. In this study, LUAD was divided into three different metabolism-relevant subtypes, and the repeatability of this subtyping was verified in the context of several test sets. Variances in metabolomic features, prognostic traits, transcriptome characteristics, immune infiltration, clinical characteristics, and drug sensitivities among the three subtypes were compared. Our results revealed that C1 possessed the most abundant metabolic pathway. Compared to C2 and C3, C1 possessed 35 different metabolic pathways, the majority of which were upregulated. Therefore, we defined C1 as the metabolically active subtype. There were five different metabolic pathways in C2, all of which were related to the lipid metabolism. Therefore, we defined C2 as the lipid metabolism-related subtype. Subtype C3 possessed five metabolic pathways. Analysis of the clinical features of the groups revealed that the majority of samples in C1 were in advanced clinical and pathological stages. Moreover, subtype C1 exhibited lower enrichment in LUAD suppressor signatures (such as HIPPO and WNT) than did the other two subtypes. Tumor microenvironment-related estimations revealed that the immune score, matrix score, and immune infiltration of C1 were significantly lower than those of the C2 and C3 subtypes. Studies have demonstrated that lower immune and matrix scores are associated with later-stage tumors and worse overall survival outcomes.<sup>41</sup> These results are consistent with those of our analysis, indicating that subtype C1 exhibited the worst prognosis in the training and testing sets. Compared to C2, which is involved in lipid metabolism, C1 was involved in a variety of metabolic processes, including amino acid, glucose, and lipid metabolism. Abundant metabolic signatures indicate that patients with C1 may benefit from metabolic therapies. In this era of increasing drug resistance, metabolic therapies for specific metabolic processes offer an alternative regimen for LUAD treatment.

Subtype C3 exhibited higher mutation rates in *TP53* than subtypes C1 and C2 did. As a well-known tumor suppressor gene, *TP53* is frequently mutated in diverse cancers and exhibits a 46% mutation rate in LUAD.<sup>42</sup> *TP53* plays an important role in many cellular processes, including cell cycle arrest, apoptosis, DNA damage response, metabolism, inflammation, immune response, angiogenesis, and metastasis.<sup>43</sup> Several studies have demonstrated that P53 plays an important role in tumor recognition by the immune system.<sup>44–52</sup> According to the heatmap presented in Figure 5(a), the total *TP53* mutation rate of the three subtypes was 53%, and the mutation rate of the C3 subtype was significantly higher than those of the other two subtypes. According to the classification of LUAD proposed by TCGA Research Network, LUAD can be divided into three subtypes: terminal respiratory unit, proximal inflammatory, and proximal proliferative transcriptional subtypes.<sup>53</sup> Subtype proximal inflammatory is characterized by a mutation in *TP53*, which is consistent with the characteristics of subtype C3 in this study. Mutations in *TP53* are associated with poorer survival of patients treated with adjuvant platinum-based therapy for completely resected stages II and III NSCLC. This finding has important implications for patients with subtype C3 in terms of therapeutic approach.

The immune checkpoint inhibitors nivolumab and pembrolizumab can improve survival outcomes in patients with lung cancer, and they have been approved by the Food and Drug Administration for the treatment of patients with advanced and recurrent lung cancer.54,55 It is important to identify objective molecular markers that can predict the effects of immune therapy. Based on ongoing clinical trials and approved immunosuppressants for specific cancer types, we selected 13 potential immune checkpoint genes. Our results revealed that the majority of the immune checkpoint genes for subtype C3 were highly expressed, indicating that patients with C3 subtype exhibited a better response to anti-PD-1 therapy. Moreover, we retrospectively analyzed prognostic and predictive markers of efficacy reported by other investigators. In recent years, neoantigens derived from oncogenic driver gene mutations have become a major focus in immunotherapy efficacy studies.<sup>56-58</sup> Typically, a high neoantigen load in tumors has been linked to an enhanced response to ICI.59,60 Our study revealed that the neoantigen loads of subtypes C3 and C1 were significantly higher than that of C2, suggesting that subtypes C1 and C3 may benefit from immunotherapy, and patients with subtype C2 are less likely to experience effective immunotherapy outcomes. Several studies have reported that tumors with CD8<sup>+</sup> T cell infiltration and high PD-L1 expression can benefit from ICI therapy.<sup>61,62</sup> Consistent with the studies by Chen and Peng, the degree of infiltration of activated natural killer cells and CD8<sup>+</sup> T cells of subtype C3 was higher than those of the other two groups,<sup>63,64</sup> suggesting that patients with subtype C3 are most likely to benefit from immunotherapy. TMB is also a key factor that affects the efficacy of immunotherapy.65-67 Our data revealed significant differences in the TMB among the subtypes. The TMB of subtype C3 was significantly higher than those of the other two subtypes. The above studies confirmed that patients with subtype C3 are most likely to benefit from immunotherapy from several different perspectives.

We further examined the  $IC_{50}$  values of 100 broadly employed drugs for LUAD therapy (primarily moleculartargeted drugs) in the three subgroups. The top 12 drugs with the largest gaps are presented in Figure 6(d). The data revealed that the  $IC_{50}$  of docetaxel in subtype C3 was significantly lower than that in C1 and C2, suggesting that different chemotherapy regimens could be considered based on metabolic subtypes. Preclinical studies have demonstrated that drugs targeting PKM2 are effective against certain types of cancer and have been established as safe for patients in early phase clinical trials.<sup>68–70</sup> We observed that shikonin, a PKM2 inhibitor, exhibited significantly different IC<sub>50</sub> values among the subtypes examined in this study. Notably, the IC<sub>50</sub> of subtype C1 was significantly higher than those of the other two subtypes, suggesting that C1 may be resistant to this drug. These findings have important implications for future clinical research and practice.

This study developed a metabolic signature that predicts the prognosis of patients with LUAD, thereby providing robust support for personalized treatment. The signature consisted of 90 metabolic genes. The prognosis of patients in the high-risk group was significantly worse than that of other patients in the training and testing sets. This metabolic signature can assist in identifying patients who may benefit from specific targeted therapies or immunotherapies and may offer new insights for the early screening of LUAD, particularly for subtypes with weak immune responses or high metabolic activity. Future public health strategies could potentially incorporate early interventions targeting these subtypes. Additionally, we further evaluated the application potential of the generated 90-gene classifier in other cancer types. The results revealed differences in expression levels among the different types of cancer. Marked differences were observed in lung squamous cell carcinoma, pancreatic cancer, cholangiocarcinoma, and gastric adenocarcinoma. This finding holds potential value for broader public health strategies addressing cancer. Overall, our work provides an in-depth analysis of LUAD subtypes based on large-scale datasets (TCGA and GEO), enhancing a better understanding of the metabolic hallmarks of LUAD and provided meaningful reference information for individualized treatment and prognosis prediction. However, this study has some limitations. First, the study primarily relies on publicly available databases, and although data processing was rigorous, it may still be affected by data acquisition standards and sample heterogeneity. More clinical and demographic characteristics of patients with LUAD should be included in our analysis to comprehensively and systematically reflect the factors influencing LUAD metabolism profiles. Second, our results must be validated using larger sample sizes and cohorts with more statistical power. Third, the identification of LUAD subtypes was based on bioinformatics analysis, which may pose certain limitations for practical clinical application. Therefore, validation of clinical samples and biological experiments are necessary to understand the differences in mechanisms among the three metabolism-relevant subtypes of LUAD. Fourth, Group C2 has a favorable prognosis, characterized by low TMB and copy-number variations, and is speculated to respond well to conventional treatments such as surgery combined with adjuvant chemotherapy. However, current clinical trial data for this subgroup are insufficient to determine whether additional targeted or immunotherapies are needed. Future research should validate the treatment requirements and potential benefits for Group C2 through prospective clinical trials. Moreover, it is essential for future studies to validate our findings in other histological types of lung cancer to determine the universality of the metabolic subtypes. Finally, owing to the retrospective nature of the cohort, a degree of selection bias was inevitable.

## Conclusions

In summary, this study classified LUAD from the perspective of metabolism and proposed three subtypes. C1 was closely related to metabolic processes and was in accordance with the characteristics of established LUAD of the proximal proliferative subtype. C2 exhibited a good prognosis similar to that of the terminal respiratory unit subtype. Subtype C3 possessed a higher level of TMB, *TP53* mutation, and neoantigens, and was most likely to benefit from immunotherapy. Due to the high accuracy of the 90-gene classifier, our study may improve the diagnosis, prognosis, and evaluation of targeted therapies for LUAD.

#### Acknowledgements

We would like to thank Editage for language editing.

#### **ORCID** iD

Leifeng Liang (D) https://orcid.org/0000-0003-1935-4100

#### **Ethical considerations**

This study did not include any studies with human participants or animals performed by any of the authors.

#### **Author contributions**

LQ and ZL: conceptualization. YW and LFL: methodology. LQ and YW: software. LL, ZL, and YW: investigation. LFL: resources. LFL, ZL, and YW: data curation. LQ and ZL: original draft preparation. LQ, ZL, YW, LFL, and LL: writing, review and editing. LFL: supervision. LFL: funding acquisition. All the authors have read and agreed to the published version of the manuscript.

#### Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by grants from the National Natural Science Foundation of China (No. 82260627) and the Guangxi Science and Technology Major Project (GuikeAA22096030).

#### Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

#### Data availability statement

Publicly available datasets were analyzed in this study. These data can be found in The Cancer Genome Atlas (https://portal.gdc.cancer.gov/) and the Gene Expression Omnibus (https://www.ncbi. nlm.nih.gov/geo/).

#### Supplemental material

Supplemental material for this article is available online.

#### References

- Cancer Genome Atlas Research Network. Comprehensive molecular profiling of lung adenocarcinoma. *Nature* 2014; 511(7511): 543–550.
- Yu KH, Berry GJ, Rubin DL, et al. Association of omics features with histopathology patterns in lung adenocarcinoma. *Cell Syst* 2017; 5(6): 620–627.e3.
- Bradley CJ, Yabroff KR, Mariotto AB, et al. Antineoplastic treatment of advanced-stage non-small-cell lung cancer: treatment, survival, and spending (2000–2011). *J Clin Oncol* 2017; 35(5): 529–535.
- Hu HC, Sun ZL, Li Y, et al. The histologic classifications of lung adenocarcinomas are discriminable by unique lineage backgrounds. *J Thorac Oncol* 2016; 11(12): 2161–2172.
- Johnson ME, Cantalupo PG and Pipas JM. Identification of head and neck cancer subtypes based on human papillomavirus presence and E2F-regulated gene expression. *mSphere* 2018; 3(1): e00580.
- Wu T, Wang YF, Jiang RH, et al. A pathways-based prediction model for classifying breast cancer subtypes. *Oncotarget* 2017; 8(35): 58809–58822.
- Hedenfalk I, Duggan D, Chen YD, et al. Gene-expression profiles in hereditary breast cancer. *N Engl J Med* 2001; 344(8): 539–548.
- Alizadeh AA, Eisen MB, Davis RE, et al. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature* 2000; 403(6769): 503–511.
- Hu FY, Zhou YX, Wang Q, et al. Gene expression classification of lung adenocarcinoma into molecular subtypes. *IEEE/* ACM Trans Comput Biol Bioinform 2020; 17(4): 1187–1197.
- 10. Pavlova NN and Thompson CB. The emerging hallmarks of cancer metabolism. *Cell Metab* 2016; 23(1): 27–47.
- Vander Heiden MG and DeBerardinis RJ. Understanding the intersections between metabolism and cancer biology. *Cell* 2017; 168(4): 657–669.
- Teicher BA, Linehan WM and Helman LJ. Targeting cancer metabolism. *Clin Cancer Res* 2012; 18(20): 5537–5545.
- 13. Luengo A, Gui DY and Vander Heiden MG. Targeting metabolism for cancer therapy. *Cell Chem Biol* 2017; 24(9): 1161–1180.
- Vernieri C, Casola S, Foiani M, et al. Targeting cancer metabolism: dietary and pharmacologic interventions. *Cancer Discov* 2016; 6(12): 1315–1333.
- Zhang M, Wang HZ, Peng RY, et al. Metabolism-associated molecular classification of colorectal cancer. *Front Oncol* 2020; 10: 602498.
- Tomczak K, Czerwinska P and Wiznerowicz M. The Cancer Genome Atlas (TCGA): an immeasurable source of knowledge. *Contemp Oncol (Pozn)* 2015; 19(1A): A68–A77.
- Barrett T, Troup DB, Wilhite SE, et al. NCBI GEO: mining tens of millions of expression profiles—database and tools update. *Nucleic Acids Res* 2007; 35: D760–D765.
- Colaprico A, Silva TC, Olsen C, et al. TCGAbiolinks: an R/ Bioconductor package for integrative analysis of TCGA data. *Nucleic Acids Res* 2016; 44(8): e71.
- 19. Rousseaux S, Debernardi A, Jacquiau B, et al. Ectopic activation of germline and placental genes identifies aggressive

metastasis-prone lung cancers. *Sci Transl Med* 2013; 5(186): 186ra66.

- Okayama H, Kohno T, Ishii Y, et al. Identification of genes upregulated in ALK-positive and EGFR/KRAS/ALK-negative lung adenocarcinomas. *Cancer Res* 2012; 72(1): 100–111.
- Yamauchi M, Yamaguchi R, Nakata A, et al. Epidermal growth factor receptor tyrosine kinase defines critical prognostic genes of stage I lung adenocarcinoma. *PLoS One* 2012; 7(9): e43923.
- Botling J, Edlund K, Lohr M, et al. Biomarker discovery in non-small cell lung cancer: integrating gene expression profiling, meta-analysis, and tissue microarray validation. *Clin Cancer Res* 2013; 19(1): 194–204.
- Jabs V, Edlund K, Konig H, et al. Integrative analysis of genomewide gene copy number changes and gene expression in nonsmall cell lung cancer. *PLoS One* 2017; 12(11): e0187246.
- Lohr M, Hellwig B, Edlund K, et al. Identification of sample annotation errors in gene expression datasets. *Arch Toxicol* 2015; 89(12): 2265–2272.
- Tang H, Xiao GH, Behrens C, et al. A 12-gene set predicts survival benefits from adjuvant chemotherapy in non-small cell lung cancer patients. *Clin Cancer Res* 2013; 19(6): 1577–1586.
- Hight SK, Mootz A, Kollipara RK, et al. An in vivo functional genomics screen of nuclear receptors and their co-regulators identifies *FOXA1* as an essential gene in lung tumorigenesis. *Neoplasia* 2020; 22(8): 294–310.
- 27. Possemato R, Marks KM, Shaul YD, et al. Functional genomics reveal that the serine synthesis pathway is essential in breast cancer. *Nature* 2011; 476(7360): 346–350.
- 28. Gaujoux R and Seoighe C. A flexible R package for nonnegative matrix factorization. *BMC Bioinform* 2010; 11: 367.
- Brunet JP, Tamayo P, Golub TR, et al. Metagenes and molecular pattern discovery using matrix factorization. *Proc Natl Acad Sci U S A* 2004; 101(12): 4164–4169.
- Hanzelmann S, Castelo R and Guinney J. GSVA: gene set variation analysis for microarray and RNA-Seq data. BMC Bioinform 2013; 14: 7.
- Rosario SR, Long MD, Affronti HC, et al. Pan-cancer analysis of transcriptional metabolic dysregulation using The Cancer Genome Atlas. *Nat Commun* 2018; 9(1): 5330.
- Sanchez-Vega F, Mina M, Armenia J, et al. Oncogenic signaling pathways in The Cancer Genome Atlas. *Cell* 2018; 173(2): 321–337.e10.
- Ritchie ME, Phipson B, Wu D, et al. Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* 2015; 43(7): e47.
- Becht E, Giraldo NA, Lacroix L, et al. Estimating the population abundance of tissue-infiltrating immune and stromal cell populations using gene expression (vol 17, 218, 2016). *Genome Biol* 2016; 17(1): 218.
- Barbie DA, Tamayo P, Boehm JS, et al. Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1. *Nature* 2009; 462(7269): 108–112.
- 36. Yoshihara K, Shahmoradgoli M, Martinez E, et al. Inferring tumor purity and stromal and immune cell admixture from expression data. *Nat Commun* 2013; 4: 2612.
- Beroukhim R, Getz G, Nghiemphu L, et al. Assessing the significance of chromosomal aberrations in cancer: methodology and application to glioma. *Proc Natl Acad Sci U S A* 2007; 104(50): 20007–20012.

- Tang ZF, Li CW, Kang BX, et al. GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses. *Nucleic Acids Res* 2017; 45(W1): W98–W102.
- Han H, Cho JW, Lee S, et al. TRRUST v2: an expanded reference database of human and mouse transcriptional regulatory interactions. *Nucleic Acids Res* 2018; 46(D1): D380–D386.
- Lu XF, Meng JL, Zhou YJ, et al. MOVICS: an R package for multi-omics integration and visualization in cancer subtyping. *Bioinformatics* 2020; 36(22–23): 5539–5541.
- Kurebayashi Y, Ojima H, Tsujikawa H, et al. Landscape of immune microenvironment in hepatocellular carcinoma and its additional impact on histological and molecular classification. *Hepatology* 2018; 68(3): 1025–1041.
- Kandoth C, McLellan MD, Vandin F, et al. Mutational landscape and significance across 12 major cancer types. *Nature* 2013; 502(7471): 333–339.
- 43. Kastenhuber ER and Lowe SW. Putting p53 in context. *Cell* 2017; 170(6): 1062–1078.
- 44. Zitvogel L and Kroemer G. A p53-regulated immune checkpoint relevant to cancer. *Science* 2015; 349(6247): 476–477.
- 45. Textor S, Fiegler N, Arnold A, et al. Human NK cells are alerted to induction of p53 in cancer cells by upregulation of the NKG2D ligands ULBP1 and ULBP2. *Cancer Res* 2011; 71(18): 5998–6009.
- 46. Shatz M, Menendez D and Resnick MA. The human TLR innate immune gene family is differentially influenced by DNA stress and p53 status in cancer cells. *Cancer Res* 2012; 72(16): 3948–3957.
- 47. Guo G, Yu M, Xiao W, et al. Local activation of p53 in the tumor microenvironment overcomes immune suppression and enhances antitumor immunity. *Cancer Res* 2017; 77(9): 2292–2305.
- Jiang ZH, Liu ZX, Li MY, et al. Immunogenomics analysis reveals that *TP53* mutations inhibit tumor immunity in gastric cancer. *Transl Oncol* 2018; 11(5): 1171–1187.
- Wang B, Niu DD, Lai LY, et al. p53 increases MHC class I expression by upregulating the endoplasmic reticulum aminopeptidase ERAP1. *Nat Commun* 2013; 4: 2359.
- Dong ZY, Zhong WZ, Zhang XC, et al. Potential predictive value of *TP53* and *KRAS* mutation status for response to PD-1 blockade immunotherapy in lung adenocarcinoma. *Clin Cancer Res* 2017; 23(12): 3012–3024.
- 51. Xiao WJ, Du N, Huang TY, et al. *TP53* mutation as potential negative predictor for response of anti-CTLA-4 therapy in metastatic melanoma. *Ebiomedicine* 2018; 32: 119–124.
- Ham SW, Jeon HY, Jin X, et al. TP53 gain-of-function mutation promotes inflammation in glioblastoma. *Cell Death Differ* 2019; 26(3): 409–425.
- Cancer Genome Atlas Research Network. Comprehensive molecular profiling of lung adenocarcinoma. *Nature* 2018; 559(7715): 543–550.
- Carbone DP, Reck M, Paz-Ares L, et al. First-line nivolumab in stage IV or recurrent non-small-cell lung cancer. N Engl J Med 2017; 376(25): 2415–2426.

- Antonia SJ, Villegas A, Daniel D, et al. Durvalumab after chemoradiotherapy in stage III non-small-cell lung cancer. *N Engl J Med* 2017; 377(20): 1919–1929.
- Bobisse S, Foukas PG, Coukos G, et al. Neoantigen-based cancer immunotherapy. *Ann Transl Med* 2016; 4(14): 262.
- Balachandran VP, Luksza M, Zhao JN, et al. Identification of unique neoantigen qualities in long-term survivors of pancreatic cancer. *Nature* 2017; 551(7681): 512–516.
- Pan D, Kobayashi A, Jiang P, et al. A major chromatin regulator determines resistance of tumor cells to T cell-mediated killing. *Science* 2018; 359(6377): 770–775.
- Matsushita H, Sato Y, Karasaki T, et al. Neoantigen load, antigen presentation machinery, and immune signatures determine prognosis in clear cell renal cell carcinoma. *Cancer Immunol Res* 2016; 4(5): 463–471.
- Howitt BE, Shukla SA, Sholl LM, et al. Association of polymerase e-mutated and microsatellite-instable endometrial cancers with neoantigen load, number of tumor-infiltrating lymphocytes, and expression of PD-1 and PD-L1. *JAMA Oncol* 2015; 1(9): 1319–1323.
- Teng MWL, Ngiow SF, Ribas A, et al. Classifying cancers based on T-cell infiltration and PD-L1. *Cancer Res* 2015; 75(11): 2139–2145.
- Ock CY, Keam B, Kim S, et al. Pan-cancer immunogenomic perspective on the tumor microenvironment based on PD-L1 and CD8 T-cell infiltration. *Clin Cancer Res* 2016; 22(9): 2261–2270.
- Peng J, Zou D, Gong WX, et al. Deep neural network classification based on somatic mutations potentially predicts clinical benefit of immune checkpoint blockade in lung adenocarcinoma. *Oncoimmunology* 2020; 9(1): 1734156.
- Chen Y, Miao SY and Zhao WC. Identification and validation of significant gene mutations to predict clinical benefit of immune checkpoint inhibitors in lung adenocarcinoma. *Am J Transl Res* 2021; 13(3): 1051–1063.
- Yarchoan M, Albacker LA, Hopkins AC, et al. PD-L1 expression and tumor mutational burden are independent biomarkers in most cancers. *JCI Insight* 2019; 4(6): e126908.
- Goodman AM, Kato S, Bazhenova L, et al. Tumor mutational burden as an independent predictor of response to immunotherapy in diverse cancers. *Mol Cancer Ther* 2017; 16(11): 2598–2608.
- Wang ZJ, Duan JC, Cai SL, et al. Assessment of blood tumor mutational burden as a potential biomarker for immunotherapy in patients with non-small cell lung cancer with use of a next-generation sequencing cancer gene panel. *JAMA Oncol* 2019; 5(5): 696–702.
- Anastasiou D, Yu YM, Israelsen WJ, et al. Pyruvate kinase M2 activators promote tetramer formation and suppress tumorigenesis. *Nat Chem Biol* 2012; 8(10): 839–847.
- Kung C, Hixon J, Kosinski PA, et al. AG-348 enhances pyruvate kinase activity in red blood cells from patients with pyruvate kinase deficiency. *Blood* 2017; 130(11): 1347–1356.
- Grace RF, Rose C, Layton DM, et al. Safety and efficacy of mitapivat in pyruvate kinase deficiency. N Engl J Med 2019; 381(10): 933–944.