Applying machine learning algorithms to develop a survival prediction model for lung adenocarcinoma based on genes related to fatty acid metabolism

Dan Cong†, Yanan Zhao†, Wenlong Zhang, Jun Li* and
Yuansong Bai*

Department of Oncology and Hematology, China-Japan Union Hospital of Jilin University, Changchun,
China

Background: The progression of lung adenocarcinoma (LUAD) may be related to abnormal fatty acid metabolism (FAM). The present study investigated the relationship between FAM-related genes and LUAD prognosis.

Methods: LUAD samples from The Cancer Genome Atlas were collected. The scores of FAM-associated pathways from the Kyoto Encyclopedia of Genes and Genomes website were calculated using the single sample gene set enrichment analysis. ConsensusClusterPlus and cumulative distribution function were used to classify molecular subtypes for LUAD. Key genes were obtained using limma package, Cox regression analysis, and six machine learning algorithms (GBM, LASSO, XGBoost, SVM, random forest, and decision trees), and a RiskScore model was established. According to the RiskScore model and clinical features, a nomogram was developed and evaluated for its prediction performance using a calibration curve. Differences in immune abnormalities among patients with different subtypes and RiskScores were analyzed by the Estimation of STromal and Immune cells in MAlignant Tumours using Expression data, CIBERSORT, and single sample gene set enrichment analysis. Patients' drug sensitivity was predicted by the pRRophetic package in R language.

Results: LUAD samples had lower scores of FAM-related pathways. Three molecular subtypes (C1, C2, and C3) were defined. Analysis on differential prognosis showed that the C1 subtype had the most favorable prognosis, followed by the C2 subtype, and the C3 subtype had the worst prognosis. The C3 subtype had lower immune infiltration. A total of 12 key genes (SLC2A1, PKP2, FAM83A, TCN1, MS4A1, CLIC6, UBE2S, RRM2, CDC45, IGF2BP1, ANGPTL4, and CD109) were screened and used to develop a RiskScore model. Survival chance of patients in the high-RiskScore group was significantly lower. The low-RiskScore group showed higher immune score and higher expression of most immune

Abbreviations: ACC2, acetyl-coA carboxylase 2; BP, biological process; CC, cellular component; CDF, cumulative distribution function; CNV, copy number variant; DCA, decision curve analysis; FABP5, fatty acid-binding protein 5; FAM, fatty acid metabolism; FASN, fatty acid synthase; GEO, Gene Expression Omnibus; GSEA, gene set enrichment analysis; GO, Gene Ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes; KM, Kaplan−Meier; LUAD, lung adenocarcinoma; MF, molecular function; SNV, single nucleotide variant; ssGSEA, single sample gene set enrichment analysis; and TCGA, The Cancer Genome Atlas.

checkpoint genes. Patients with a high RiskScore were more likely to benefit from the six anticancer drugs we screened in this study.

Conclusion: We developed a RiskScore model using FAM-related genes to help predict LUAD prognosis and develop new targeted drugs.

1 Introduction

Lung adenocarcinoma (LUAD) accounts for about 40% of primary lung tumors. LUAD is one of the tumor types that have rapid metastasis and high mortality, with a survival time shorter than 5 years (Denisenko et al., 2018; Hutchinson et al., 2019; Shi et al., 2022). LUAD, at an early stage, usually has no obvious clinical symptoms and is often diagnosed by adjuvant methods at the middle and late stages or when metastasis occurs (Patz et al., 2014; Skřičková et al., 2018; Sung et al., 2021). Although significant advances have been made in the research and clinical treatment of LUAD, the prognosis of LUAD remains dismal, despite the clinical use of chemoradiotherapy, targeted therapy, and immunotherapy. Currently, the underlying cellular and molecular mechanisms of tumor behavior remain unclear (Lin et al., 2021; Chen et al., 2022a). Therefore, molecular characteristics of LUAD should be

comprehensively investigated to improve clinical therapies and the accuracy of prognosis prediction for LUAD.

Lipid metabolism is an important metabolic process for cells. Abnormal fatty acid metabolism (FAM) in cancer cells has been increasingly studied. Carcinogenesis mechanisms of various cancers vary greatly, but they often show similar abnormalities in metabolism. Reprogramming the metabolism of glucose, fatty acids, and other biomolecules could promote the progression of tumor cells (Li and Zhang, 2016). Growing evidence demonstrated that some changes occur in tumor tissues in different processes of FAM (Amiri et al., 2018), including in deciding the types, abundance, and mechanisms of action of lipid-signaling molecules with regulatory functions (Santos and Schulze, 2012). Changes of FAM also affect the proliferation, differentiation, and metastasis of tumor cells (Yu et al., 2018). However, the characteristics and functions of genes related to FAM in LUAD have not been fully explored.
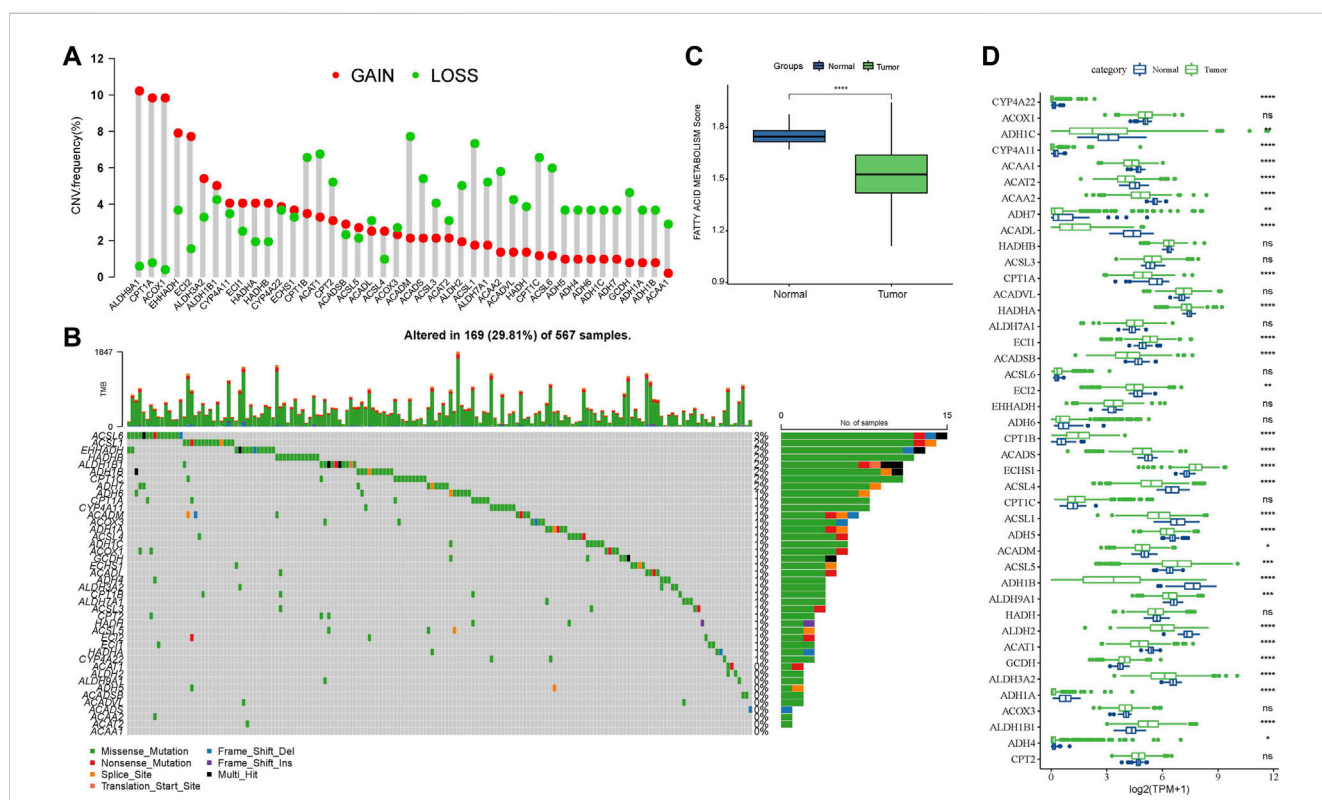
FIGURE 1
Abnormal FAM-related pathway genes in LUAD. **(A)** Proportion of gene deletion and amplification in the CNV of FAM-related pathway genes. **(B)** Waterfall diagram of mutant information in the SNV of FAM-related pathway genes. **(C)** Comparison of FAM pathway scores in LUAD and para-cancer tissues. **(D)** Comparison of expression of FAM-related pathway genes in LUAD and para-cancer tissues.
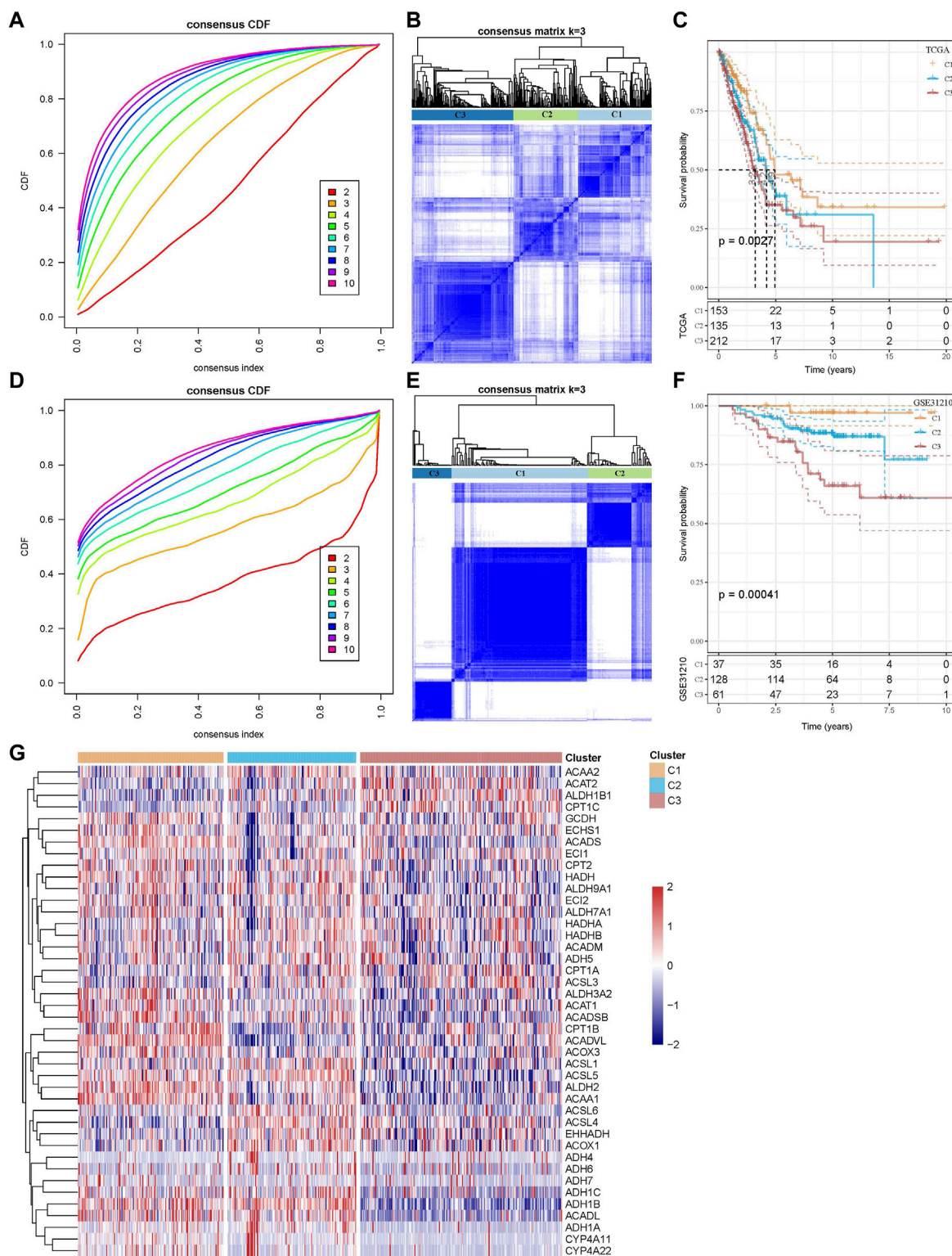
FIGURE 2
Construction of molecular subtypes based on genes related to FAM pathways. **(A)** CDF curve of TCGA cohort samples. **(B)** Clustering heatmap of samples in the TCGA cohort when consensus k = 3. **(C)** Relationship between the prognoses of three TCGA subtypes is shown by the K−M curve. **(D)** CDF curve of GSE31210 cohort samples. **(E)** Clustering heatmap of samples with consensus k = 3 in the GSE31210 cohort. **(F)** K−M curve of the relationship between the prognoses of three subtypes of GSE31210. **(G)** Heatmap of the expression of FAM-related pathway genes between three subtypes in the TCGA cohort.

FIGURE 3
Differences in clinical characteristics among molecular subtypes. **(A)** Bar chart of pair-to-pair comparison of three molecular subtypes in the TCGA cohort. **(B)** Comparative table of each clinical characteristic molecular subtype.
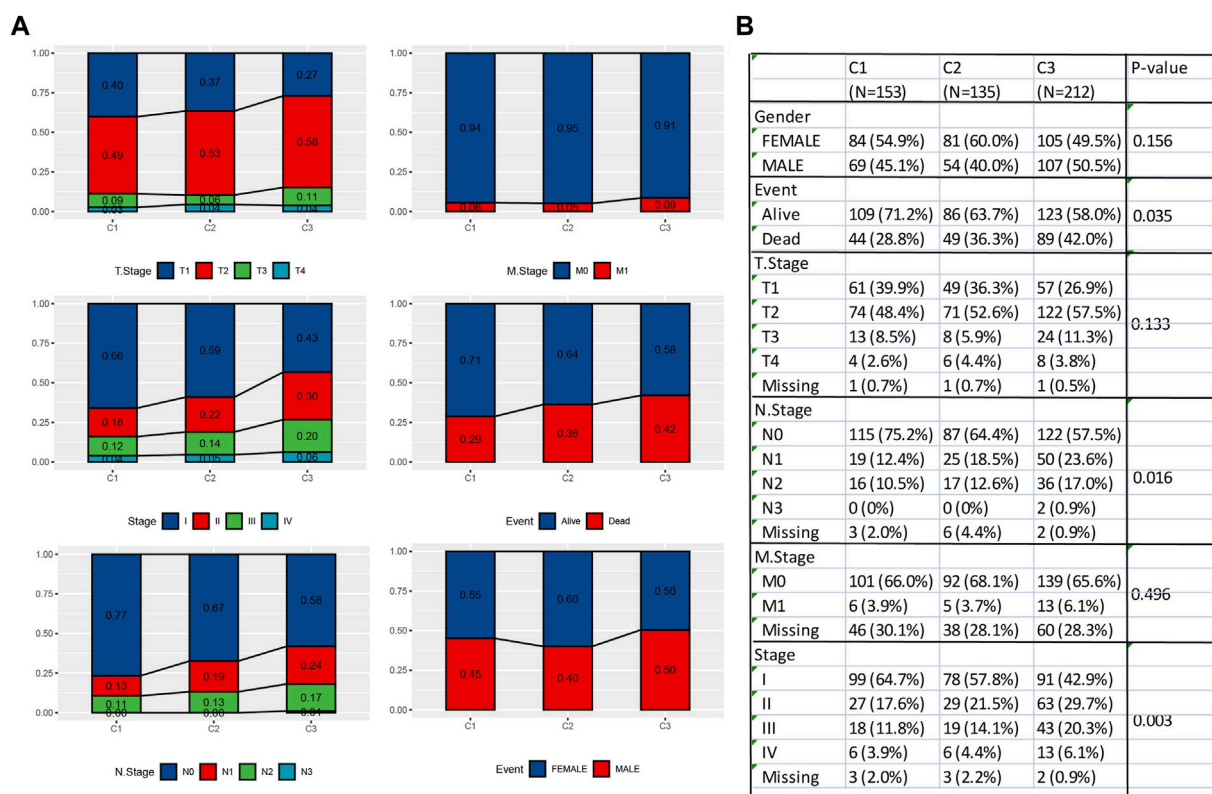
A previous study investigating abnormal FAM showed that the overexpression of fatty acid-binding protein 5 (FABP5) is related to the poor prognosis in LUAD and may be a new clinical target to treat LUAD (Garcia et al., 2022). The downregulation of fatty acid synthase (FASN) interferes with the progression of LUAD through regulating the glucose metabolism and inhibiting the AKT/ERK pathway (Chang et al., 2019). Related drugs could act on the FAM process in LUAD. For example, anlotinib controls LUAD progression through inhibiting FASN-mediated FAM (Shen et al., 2022). Chaoyang Liang et al. also showed that the overexpression of genes related to FAM enzymes (ACOT11) regulates the growth, differentiation, and metastasis of LUAD cells through a variety of signaling pathways (Liang et al., 2020). In addition, Wang et al. developed a fatty acid-related RiskScore model to predict the prognosis of lung cancer patients and identified 38 fatty acid-related genes. Among these 38 genes, eight genes (HGNAT, MCTP2, ENPP5, PLEKHA6, ANKRD29, CNTNAP2, SLC4A5, and ZNF738) have not been reported in previous lung cancer-related studies (Wang et al., 2022a). Therefore, the identification and verification of genes related to FAM may have great potential for developing new prognostic models and improving clinical treatment for LUAD.

In this study, we downloaded genomic information about the clinical characteristics of LUAD from The Cancer Genome Atlas (TCGA) and Gene Expression Omnibus (GEO) databases. Molecular subtypes related to FAM pathways were developed for LUAD, and we further established a risk assessment model based on

FAM-related genes using six machine learning algorithms. In addition, we assessed the level of immune cell infiltration and sensitivity to common drugs in different risk groups. The current study provided a better understanding of the mechanism of abnormal FAM in LUAD cells, helping to improve the therapeutic strategies for treating LUAD patients.

2 Materials and methods

2.1 Data downloading and preprocessing

2.1.1 TCGA-LUAD dataset downloading and preprocessing

Data with clinical phenotypes were obtained from the TCGA database (Liu et al., 2020). Samples without the survival time or state were eliminated to ensure that the survival time of all the included samples was longer than 0 days. Finally, 500 tumor samples and 59 para-cancer tissues samples from the TCGA dataset were obtained.

2.1.2 GEO data download and preprocessing

A set of chip data was obtained from the GEO (Barrett et al., 2013), and the probe was converted into symbol according to the annotation file. Normal tissue samples or those without clinical information were excluded to ensure that the survival time of all the included samples was longer than 0 days, and only LUAD samples
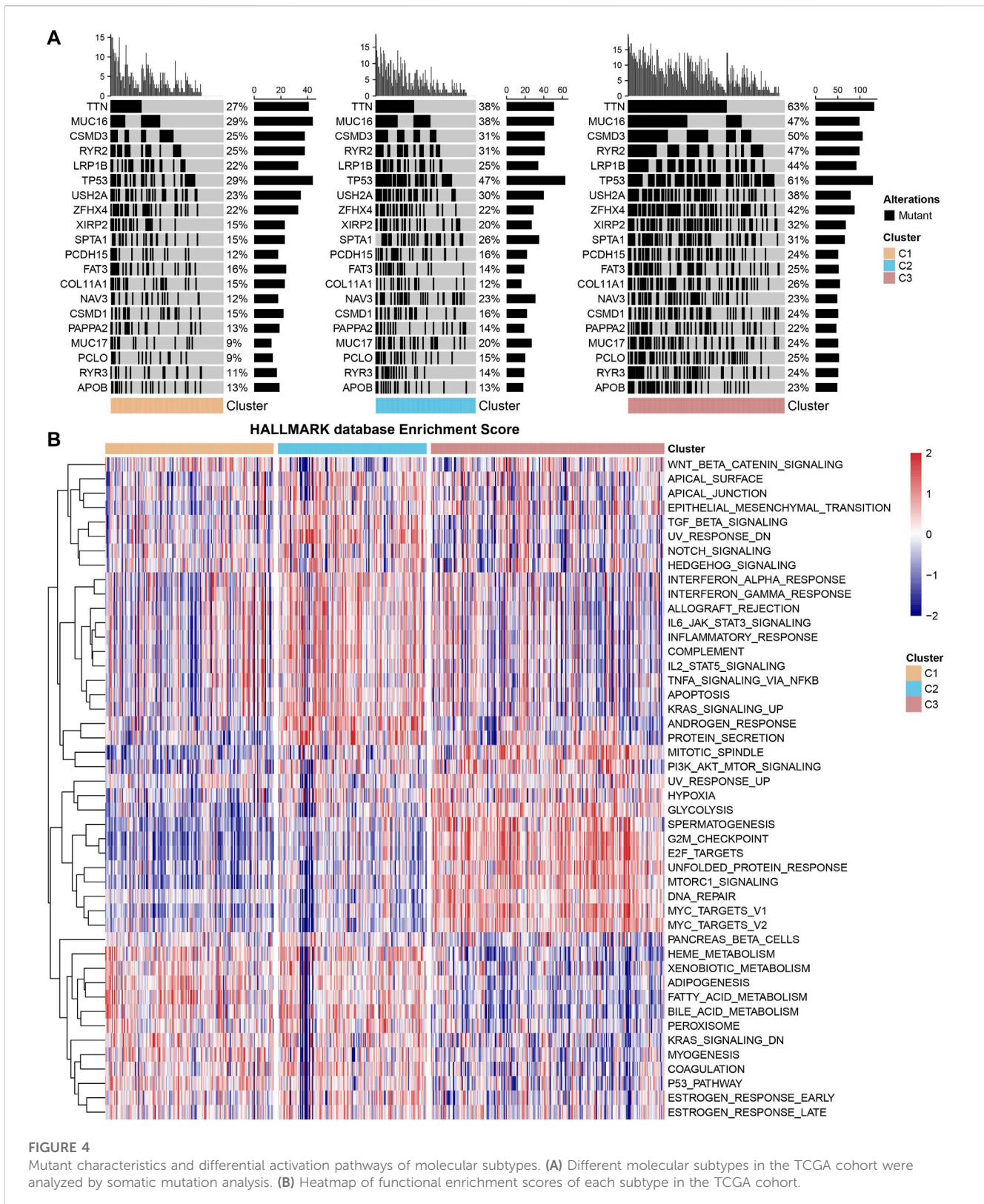
FIGURE 4
Mutant characteristics and differential activation pathways of molecular subtypes. **(A)** Different molecular subtypes in the TCGA cohort were analyzed by somatic mutation analysis. **(B)** Heatmap of functional enrichment scores of each subtype in the TCGA cohort.

were retained through data filtering. Specifically, 289 samples were obtained from the GSE30219 dataset; 226 samples were from the GSE31210 dataset; 196 samples were from the GSE37745 dataset; and 127 samples were from the GSE50081 dataset.

2.1.3 Acquisition of FAM-related genes

From the Kyoto Encyclopedia of Genes and Genomes (KEGG), a collection of 42 related genes was downloaded (Kanehisa and Goto, 2000).

FIGURE 5
Analysis of immune abnormalities of molecular subtypes and enrichment analysis of differential genes. **(A)** Comparison of immune scores among different subtypes. **(B)** Comparison of 22 immune cell scores among different subtypes. **(C)** Comparison of the abundance of 28 immune cells among different subtypes; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$, and ns represents $p > 0.05$. **(D)** Circle diagram of the top five functional enrichment analyses of differential genes GO enrichment analysis. **(E)** Circle diagram of the top 15 functional enrichment analyses for KEGG enrichment analysis of differential genes.

2.2 Classification of molecular subtypes

The ConsensusClusterPlus package was used to cluster the TCGA-LUAD and GSE31210 dataset, and the clustering heatmap of the samples was drawn (Liu et al., 2022a). Cumulative distribution function (CDF) was used to obtain the optimal clustering number and relatively stable clustering results. Three molecular subtypes (C1, C2, and C3) were then identified. To further analyze the prognosis of different molecular subtypes, Kaplan–Meier (KM) curves were drawn using the survminer package.

2.3 Filtering of differentially expressed genes and enrichment analysis

In order to further screen gene sets related to FAM subtypes, we used the limma package to analyze the differences between C1 and C2+C3, C2 and C1+C3, and C3 and C1+C2 in the TCGA-LUAD dataset under the threshold of │ log2 (a Fold Change) │ > 1, FDR <0.05 (Ritchie et al., 2015). GO and KEGG enrichment analysis were performed on genes showing abnormal expression using the clusterProfiler software package (Yang et al., 2023).

2.4 Construction of the RiskScore model

We used univariate Cox analysis for analyzing the differentially expressed genes (DEGs) (Peng et al., 2021). The prognostic genes with $p < 0.001$ were screened. Machine learning models can be widely used in the medical field due to their excellent performance in predicting classification problems (Choi et al., 2020). Therefore, based on six machine learning algorithms, namely, GBM (Dash

et al., 2022), LASSO (Kang et al., 2021), XGBoost (Li et al., 2022a), SVM (Zhou, 2022), random forest (Utkin and Konstantinov, 2022), and decision trees (Streeb et al., 2022), the DEGs in the comparison pairs of C1 and C2+C3, C2 and C1+C3, and C3 and C1+C2 were comprehensively analyzed, and the characteristic genes were obtained by overlapping analysis. A stepwise regression method was used to further compress the characteristic genes. We calculated the β value by multivariate Cox analysis (Zhang et al., 2021a). The calculation formula of the model is as follows:

$$\text{RiskScore} = \Sigma\beta i \times \text{Exp}i.$$

In the formula, *Expi* is the expression value of key FAM-related genes and β is the Cox regression coefficient of the key genes.

According to the abovementioned formula, the RiskScore of each TCGA-LUAD sample was determined and then processed by the Z-score (DeVore, 2017). Then, the samples with a RiskScore less than 0 were categorized as the low-RiskScore group, while those with a RiskScore greater than 0 were categorized as the high-RiskScore group. Five sets of chip data (GSE31210 cohort, GSE19188 cohort, GSE30219 cohort, GSE37745 cohort, and GSE50081 cohort) were used to calculate the RiskScore by the same method. A receiver operating characteristic (ROC) curve was obtained using the timeROC package (Lu et al., 2022a).

2.5 Comparison of clinical features

The clinicopathological features (Gender, Event, T. Stage, M. Stage, N. Stage, and Stage) in different molecular subtypes and different RiskScore groups in the TCGA cohort were analyzed. The pheatmap package of R software was applied to plot a heatmap to

FIGURE 6
Key characteristic genes screening and K−M curves of six datasets. **(A)** Scatter plot of univariate Cox analysis of 493 genes associated with subtypes of FAM. **(B)** Venn diagram comparing C1 and C2 + C3 for characteristic gene screening of six algorithms. **(C)** Venn diagram comparing C2 and C1 + C3 for characteristic gene screening of six algorithms. **(D)** Venn diagram comparing C3 and C1 + C2 for characteristic gene screening of six algorithms. **(E)** K−M curve of the RiskScore model developed by the 12 genes in the TCGA cohort. **(F)** K−M curve of the RiskScore model developed by the 12 genes in the GSE31210 cohort. **(G)** K−M curve of the RiskScore model developed by the 12 genes in the GSE19188 cohort. **(H)** K−M curve of the RiskScore model developed by the 12 genes in the GSE30219 cohort. **(I)** K−M curve of the RiskScore model developed by the 12 genes in the GSE37745 cohort. **(J)** K−M curve of the RiskScore model developed by the 12 genes in the GSE50081 cohort. **(K)** Line graph of AUC for 1−5 years of RiskScore for six datasets.

examine the distribution of samples with different clinical features (Zhang et al., 2021b).
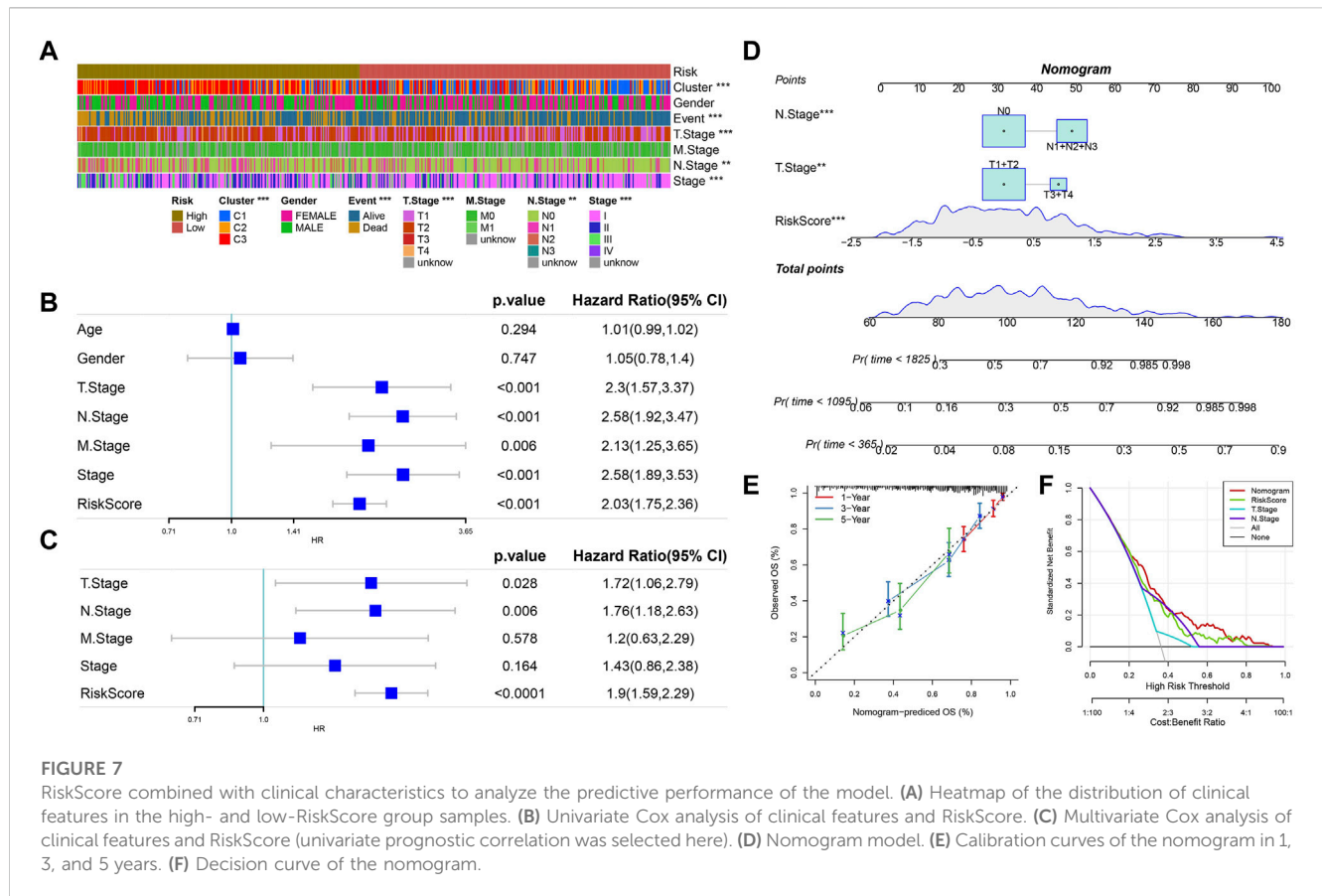
2.6 Establishment of a nomogram

The relationships between clinical features, RiskScore, and prognosis were assessed applying univariate and multivariate Cox analyses (van de Vijver et al., 2002). The model prediction efficiency was evaluated by developing a nomogram combining key clinicopathological features using the rms package (Liu et al., 2021a). A calibration curve was used to evaluate the predictive power of the model and to test the prediction performance of the nomogram (Van Calster et al., 2019). We also used the ggDCA

package to assess the stability of the decision curve analysis (DCA) and to plot the calibration curve and DCA for the nomogram in predicting 1-, 3-, and 5-year prognosis (Van Calster et al., 2018).

2.7 Mutation analysis

Data for copy number variant (CNV) were downloaded to compare the deletion or amplification of genes associated with FAM pathways (Wu et al., 2020). Then, mutation data of single nucleotide variants (SNVs) were downloaded and a waterfall map was generated using the maftools package to display SNV mutations in FAM-related pathway genes (Li et al., 2022b).

FIGURE 7
RiskScore combined with clinical characteristics to analyze the predictive performance of the model. **(A)** Heatmap of the distribution of clinical features in the high- and low-RiskScore group samples. **(B)** Univariate Cox analysis of clinical features and RiskScore. **(C)** Multivariate Cox analysis of clinical features and RiskScore (univariate prognostic correlation was selected here). **(D)** Nomogram model. **(E)** Calibration curves of the nomogram in 1, 3, and 5 years. **(F)** Decision curve of the nomogram.

Differences in genomic changes were examined in different molecular subtypes. Mutated datasets were processed using mutect2 software (Pei et al., 2021). Genes showing a mutation frequency greater than 3 were filtered. Fisher's test was applied to detect frequently mutated genes in each subtype ($p < 0.05$).

2.8 Pathway difference analysis

The FAM pathway scores were calculated by single sample gene set enrichment analysis (ssGSEA) (Zhuang et al., 2021). Differences in FAM-related pathway scores between LUAD and para-cancer tissues were compared by the Wilcoxon signed-rank test (Divine et al., 2013). We used the pheatmap package to draw heatmap to show the expression of related genes (Zheng et al., 2022).

In order to characterize biological process pathways, the GSVA software package (Hänzelmann et al., 2013) was used to analyze all the relevant gene sets in the Hallmark database. The Kruskal test was performed to examine differentially activated pathways in different molecular subtypes (Liu et al., 2021b). Significant pathways were selected under $p < 0.05$, and the heatmap of functional enrichment scores of each subtype was generated.

2.9 Comparison of immune abnormalities

Immune infiltration was evaluated by Estimation of STromal and Immune cells in MAlignant Tumors using Expression (ESTIMATE), and the differences in immune scores were compared (Yang et al., 2021). Then, CIBERSORT algorithm was used to calculate the abundance of 22 kinds of immune cells and compare the differences in immune cell scores (Zhang et al., 2022a). A variety of immune cell characteristic genes were identified (Charoentong et al., 2017). We compared the differences of 28 immune cell scores using the ssGSEA. Furthermore, the expression of gene multiple immune checkpoint genes was analyzed between different RiskScore groups (Danilova et al., 2019).

2.10 Drug sensitivity analysis

The R language pRRophetic package could be used to predict patient sensitivity to drugs. Several commonly used drugs such as erlotinib, paclitaxel, MG-132, rapamycin, sunitinib, and cisplatin were selected (Skalniak et al., 2013; Landi and Cappuzzo, 2015; Bhaoighill and Dunlop, 2019; Wang et al., 2020; Lu et al., 2022b).

2.11 Statistical analysis

This study mainly used R software for statistical analysis. A $p < 0.05$ was defined as a statistically significant difference. The Wilcoxon test was used to assess differences in immune abnormalities between RiskScore groups.
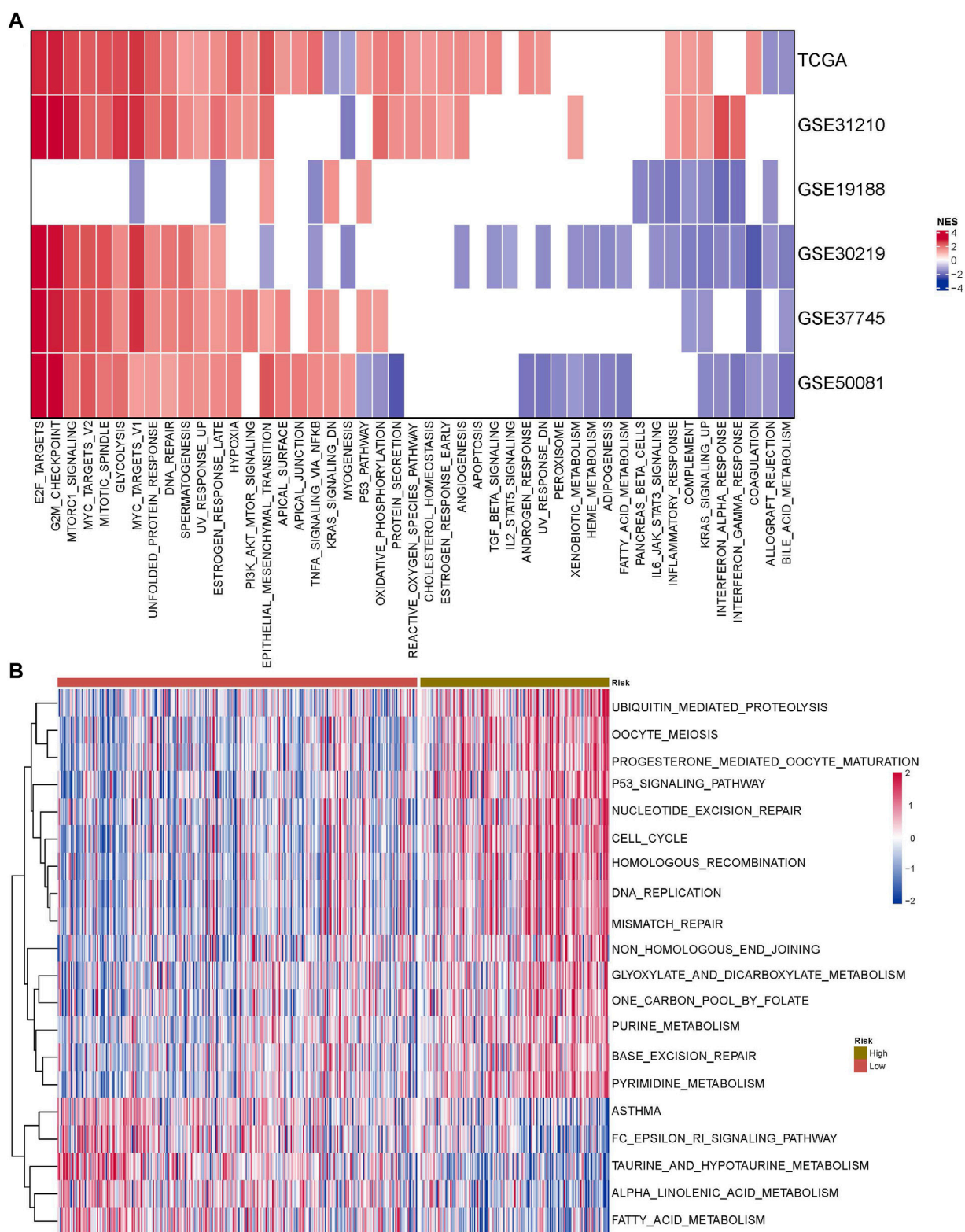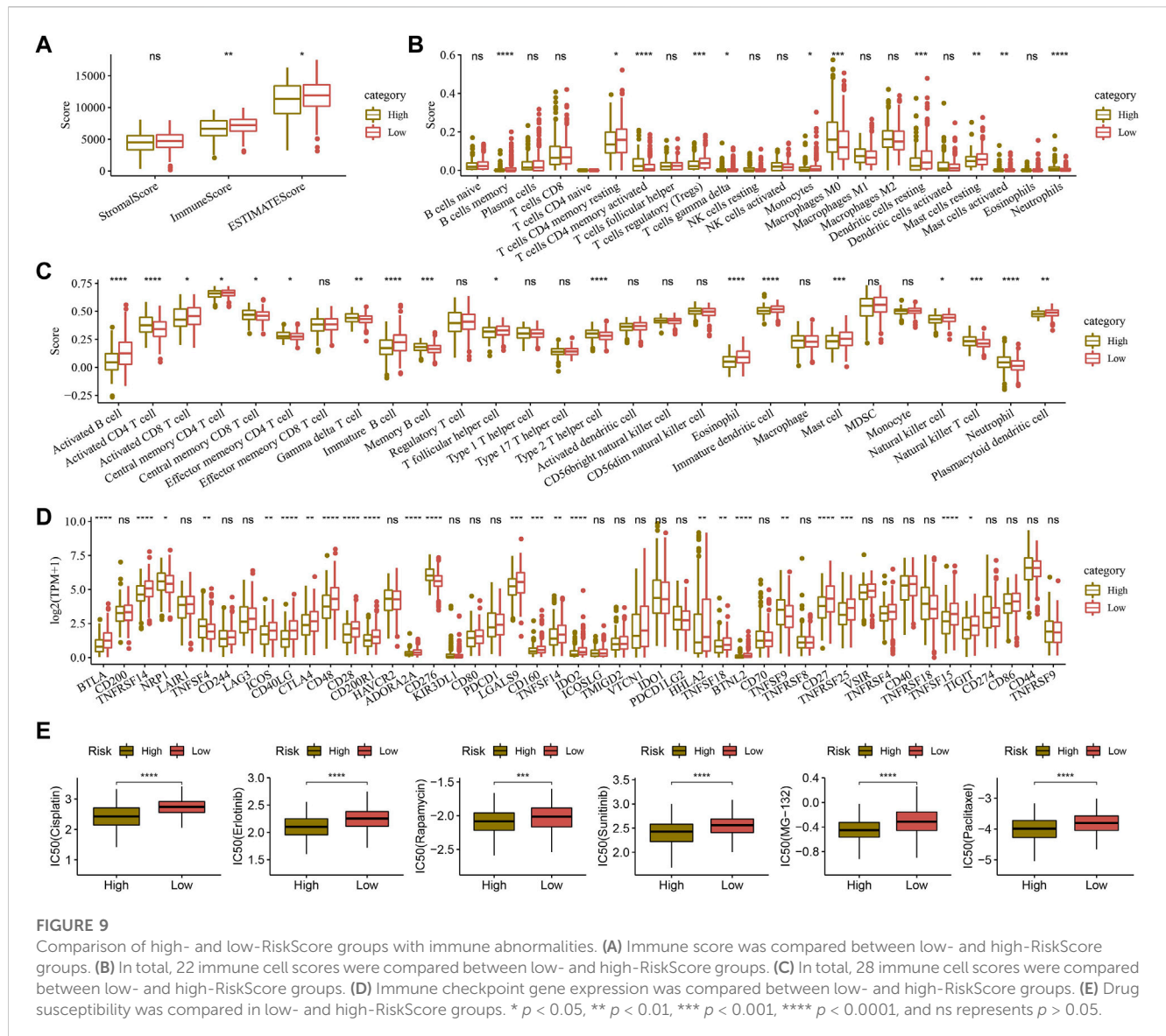
FIGURE 8
Combined with the RiskScore model, potential regulatory pathways were identified. **(A)** Pathway enrichment score heatmap obtained by six datasets in the HALLMARK gene set. **(B)** Heatmap of correlated pathway scores in the KEGG database TCGA cohort in high- and low-RiskScore groups.

FIGURE 9
Comparison of high- and low-RiskScore groups with immune abnormalities. **(A)** Immune score was compared between low- and high-RiskScore groups. **(B)** In total, 22 immune cell scores were compared between low- and high-RiskScore groups. **(C)** In total, 28 immune cell scores were compared between low- and high-RiskScore groups. **(D)** Immune checkpoint gene expression was compared between low- and high-RiskScore groups. **(E)** Drug susceptibility was compared in low- and high-RiskScore groups. * *p* < 0.05, ** *p* < 0.01, *** *p* < 0.001, **** *p* < 0.0001, and ns represents *p* > 0.05.

3 Results

3.1 Abnormal FAM-related pathway genes in LUAD

After selecting FAM-related pathway genes, we found that some genes, such as ACADM, ACSL1, and CPT1C, tended to show deletion, while some genes, such as ALDH9A1, CPT1A, and ACOX1, tended to show amplification (Figure 1A). The SNV of the genes related to FAM pathways was shown in waterfall diagram, and it was found that ACSL6 had the highest mutation rate and was mostly missense mutation (Figure 1B). A comparison of scores of FAM pathways showed that LUAD tissues had a lower score of FAM pathways compared to para-cancer tissues (Figure 1C). The expression of 24 FAM-related pathway genes was more active in para-cancer tissues than that in LUAD tissues, such as ACAA1, ACAT2, and ADH1B (Figure 1D). These results suggested that FAM-related genes may have an impact on the progression of LUAD.

3.2 Classification of molecular subtypes based on genes related to FAM pathways

In the TCGA dataset, the CDF curve showed that cluster 3 was a relatively stable clustering (Figure 2A). Finally, three molecular subtypes of C1, C2, and C3 were defined based on the sample clustering heatmap (Figure 2B). Further analysis of the K–M curves for the three molecular subtypes showed significant differences in terms of prognostic survival among the three subtypes (*p* = 0.0027). Overall, C1 had the best survival outcome, followed by C2 and C3 (Figure 2C).

We used the abovementioned methods to analyze and classify the GSE31210 dataset. The CDF results showed that cluster 3 also had relatively stable clustering results (Figure 2D). At k = 3, the three molecular subtypes were significantly different (Figure 2E). There were also significant differences in prognostic survival among the three subtypes (*p* = 0.00041). The survival of C1 was found to be the most favorable, while that of C3 was the worst, and the overall results were similar to those of the TCGA dataset (Figure 2F).

The gene expression of FAM-related pathways in the three subtypes was shown in the heatmap. It was found that the gene expression of FAM-related pathways in C1 was relatively active, while that in C3 was relatively poor (Figure 2G). These results suggested that FAM-related subtypes were associated with different prognosis, and that tumors of different subtypes had large differences in the FAM status.

3.3 Significant differences in clinicopathological characteristics among the three subtypes

The clinicopathological characteristics of different subtypes in the TCGA cohort were analyzed. There were significant differences in three clinical indicators (Event, N Stage, and Stage) of the three subtypes ($p < 0.05$). Event showed that C1 had a significantly higher survival probability compared to C2 and C3, while C3 had the worst survival outcomes. As for N stage, C1 had the highest proportion of N0, while C3 had the highest proportion of N2. In Stage, C1 had the largest proportion in stage I, while C3 had the largest proportion in stage IV. Therefore, the outcome, clinical grade, and staging of C1 were relatively favorable, while those of C3 were unfavorable (Figure 3).

3.4 Mutant characteristics and differential activation pathways of the three subtypes

Differences in genomic changes among subtypes in the TCGA cohort were analyzed. The mutation dataset of TCGA was processed by mutect2 software, and a total of 9,922 genes were screened. Fisher's test was used for screening with $p < 0.05$, which filtered 770 genes. The top 20 genes were selected for further analysis on the characteristics of somatic mutations. The results showed that C1 had the lowest mutation rate and C3 had the highest mutation rate (Figure 4A). Moreover, whether differentially activated pathways were present in different subtypes were explored. Some screened pathways were found to be significantly differentially activated in different subtypes, for example, PI3K AKT MTOR SIGNALING, G2M CHECKPOINT, and FATTY ACID METABOLISM. The FATTY ACID METABOLISM pathway was actively expressed in the C1 and C2 subtypes but less expressed in the C3 subtype (Figure 4B). This could explain a poorer prognosis of C3.

3.5 Analysis of immune abnormalities of subtypes

Analysis on differences in the immune microenvironment among the three subtypes showed that C3, with a poor prognosis, had the lowest scores of StromalScore, ImmuneScore, and ESTIMATEScore, which indicated lower immune infiltration of C3 (Figure 5A). The abundance of 18 kinds of immune cells, such as B-cell memory, was different among the three subtypes (Figure 5B). The calculation results of 28 immune cell scores, such as activated B cells and activated CD4 T cells, demonstrated differences in 26 immune cell scores among the three subtypes (Figure 5C).

3.6 Screening and enrichment of FAM-related genes

To screen gene sets associated with FAM subtypes, differential analysis was performed for C1 and C1+C2, C2 and C1+C3, and C3 and C1+C2. A total of 124 upregulated genes and 123 downregulated genes were screened in the comparison between C1 and C2+C3; 59 upregulated genes and four downregulated genes were screened in the comparison between C2 and C1+C3; and 160 upregulated genes and 276 downregulated genes were screened in the comparison between C3 and C1+C2. There were 493 DEGs in total. Circle diagrams of GO enrichment analysis on the top five functional enrichment analyses were plotted, and we observed that mitotic sister chromatid segregation was the most significant biological process. The condensed chromosome centromeric region was the most prominent cellular component. The MHC class II receptor activity was the most active molecular function (Figure 5D). Asthma was found to be the most significant pathway in the circle diagram of the top 15 functional enrichment analyses on the differential genes (Figure 5E).

3.7 Construction and evaluation of a RiskScore model

Building upon these findings, univariate Cox analysis was used to perform prognostic analysis on 493 genes related to FAM subtypes, and 143 were screened to be prognostic genes relevant to LUAD ($p < 0.001$) (Figure 6A). Subsequently, we determined feature genes using six machine algorithms (including LASSO, GBM, random forest, SVN, XGBoost, and decision trees) for C1 and C2+C3, C2 and C1+C3, and C3 and C1+C2. Through the Venn diagram, we found a total of 18 characterized genes between C1 and C2+C3 (Figure 6B), a total of 16 genes between C2 and C1+C3 (Figure 6C), and a total of 12 genes between C3 and C1+C2 (Figure 6D). Further comparison and screening showed 34 important genes for subsequent studies. Finally, the number of important genes was reduced to 12 key genes (SLC2A1, PKP2, FAM83A, TCN1, MS4A1, CLIC6, UBE2S, RRM2, CDC45, IGF2BP1, ANGPTL4, and CD109) by the stepwise regression method.

The calculation formula is as follows:

RiskScore = −0.293*SLC2A1+0.145*PKP2+0.113*FAM83A+ 0.092*TCN1-0.142*MS4A1-0.081*CLIC6+0.24*UBE2S+0.217*RRM2-0.286*CDC45+0.162*IGF2BP1+0.093*ANGPTL4+0.112*CD109.

The RiskScore of the sample in TCGA was calculated by the abovementioned model formula, and high- and low-RiskScore groups were classified. The K–M curve results showed that the survival probability of the high-RiskScore group was lower ($p < 0.0001$, Figure 6E). At the same time, five sets of chip data (GSE31210, GSE19188, GSE30219, GSE37745, and GSE50081) were used to draw K–M curves, and the survival probability of the high-RiskScore group was still lower (Figures 6F–J). The AUC values of the 1-, 3-, and 5-year RiskScore of six datasets were observed, and it was found that the values of six datasets were all around 0.7 and that the AUC values of three datasets were

consistently above 0.7, indicating that the model had a strong predictive performance (Figure 6K).

3.8 Testing the predictive performance of the RiskScore model combined with clinical characteristics

Combined with the RiskScore, the distribution of samples with multiple clinical characteristics was presented in the form of a heatmap. The results showed that the distribution of five clinical characteristics (Cluster, Event, T Stage, N Stage, and Stage) was closely correlated with that of the RiskScore (Figure 7A). Meanwhile, the univariate Cox analysis between each clinical feature and the RiskScore showed that the p-values of T Stage, N Stage, Stage, and RiskScore were all less than 0.001. The multivariate Cox analysis showed that the p-values of RiskScore, T Stage, and N Stage were all less than 0.05. Therefore, T Stage, N Stage, and RiskScore were independent prognostic factors (Figures 7B, C). A nomogram was established by the abovementioned factors. According to the results, the RiskScore had the strongest survival prediction ability (Figure 7D). The slope and distance between all the calibration curves and standard ones were similar, which verified the nomogram's prediction capability (Figure 7E). The benefit rates of the RiskScore and nomogram were significantly higher than the extremum curves, which proved that the nomogram and RiskScore had the greatest power for survival prediction (Figure 7F).

3.9 Potential regulatory pathways identified by the RiskScore model

The HALLMARK gene set was enriched in six datasets by GSEA, and it was found that E2F target and G2M checkpoint pathways had the highest scores, while bile acid metabolism and other pathways had lower scores. The FAM pathway score was also generally low (Figure 8A). We compared the predicted pathway scores of RiskScore groups in the TCGA cohort and observed that the scores of 15 pathways including ubiquitin-mediated proteolysis were higher in the high-RiskScore group, while the scores of five pathways including FAM were lower in the high-RiskScore group (Figure 8B).

3.10 Immune status and immunotherapy preference predicted by the RiskScore

The ImmuneScore and ESTIMATEScore were higher in the high-RiskScore group compared to those in the low-RiskScore group (Figure 9A). Among the 22 immune cell scores predicted, 11 immune cell scores, such as T-cell CD4 memory resting, showed significant differences between the two RiskScore groups (Figure 9B). Among the 28 immune cell scores predicted, 18 immune cell scores, such as activated CD4 T cells, showed differences between the two RiskScore groups (Figure 9C). The expression of 48 different immune

checkpoint genes was compared between the two RiskScore groups, and higher expression of 19 immune checkpoint genes, such as BTLA, TNFRSF14, ICOS, and CD48, was found in the low-RiskScore group (Figure 9D).

Cisplatin, erlotinib, rapamycin, sunitinib, MG-132, and paclitaxel were all found to be more sensitive to the high-RiskScore group, suggesting that patients in high-RiskScore groups might respond better to these drugs (Figure 9E).

4 Discussion

Lung cancer is one of the most deadly malignant tumors worldwide (Wang et al., 2022b). The metabolic reprogramming of cancer cells, particularly the modification of FAM, is firmly connected with tumor growth (Maan et al., 2018). Abnormal FAM is associated with the growth, differentiation, and metastasis of LUAD cells. Acetyl-coA carboxylase 2 (ACC2) is a key FAM enzyme. Fei-Yuan Yu et al. found that ACC2 is low-expressed in tumor cells, and its expression is negatively correlated with tumor progression (Yu et al., 2022). FASN is a homodimeric multienzymatic protein that inhibits and blocks the adipogenic pathway and hinders fatty acid synthesis. This causes apoptosis in tumor cells to overexpress FASN without affecting non-malignant cells (Relat et al., 2012). Recent studies showed that FASN expression is upregulated and overactivated in LUAD, which may be related to the progression of LUAD (Relat et al., 2012). Drug targeting FAM pathways in LUAD has been designed. For example, AZ12756122, a novel FASN inhibitor, can induce cell apoptosis, downregulate FASN expression and activity, and reduce EGFR and Akt/mTOR pathway activation (Polonio-Alcalá et al., 2022). Although the relationship between the gene expression of FAM pathways and the prognosis of LUAD has been explored, this study introduced a variety of machine learning analysis algorithms to more comprehensively analyze the genetic characteristics of FAM pathways (Ganggayah et al., 2019), which can help establish a more effective risk prediction model for LUAD based on the FAM pathway genes.

Using LUAD data from the TCGA dataset, GEO dataset, and FAM-related gene sets obtained by KEGG analysis, we found that LUAD had lower scores of FAM-related pathways. Molecular subtypes were classified using the genes related to FAM pathways, and six machine learning methods were applied to select key genes related to the three LUAD subtypes. A total of 12 key genes (SLC2A1, PKP2, FAM83A, TCN1, MS4A1, CLIC6, UBE2S, RRM2, CDC45, IGF2BP1, ANGPTL4, and CD109) were determined to be closely related to LUAD prognosis. The downregulation of SLC2A1-AS1 can inhibit LUAD cell growth and expansion, and its overexpression increases tumor cell proliferation and differentiation. PKP2 promotes the growth, division, and migration of cancer cells through activating the EGFR signaling pathway in LUAD cells (Hao et al., 2019). FAM83A-AS1 knockdown can suppress the proliferation of LUAD cells, can inhibit the expression of HIF-1α and glycolytic genes, and also plays a role in FAM (Chen et al., 2022b). High expression of TCN1, a vitamin B12-binding protein, is positively associated with cancer aggressiveness and a poor prognosis (Li et al.,

2022c). The expression level of MS4A1 in colorectal cancer is positively correlated with patients' prognosis. CLIC6 is upregulated in most obese patients with endometrial cancer (López-Ozuna et al., 2021; Mudd et al., 2021). Mengjun Zhang et al. observed that UBE2S can promote PI3K or mTOR signaling pathway, block the regulation of cell cycle, inhibit cell apoptosis, and promote the proliferation, migration, and prognosis of ovarian cancer (Zhang et al., 2022b). RRM2 is upregulated in LUAD, and high RRM2 expression is associated with a poorer survival and lower immune infiltration (Ma et al., 2020). In addition, Zhou et al. demonstrated that RRM2 is overexpressed in the cell lines and clinical samples of bladder cancer and that blocking RRM2 inhibits the growth and proliferation of cancer cells (Zhou et al., 2022). CDC45, a key protein involved in the initiation of DNA replication, is upregulated in many cancers, and its expression is significantly negatively correlated with patient prognosis (Lu et al., 2022c). JinFeng Liu et al. found that IGF2BP1 is significantly abnormally expressed in LUAD samples. Moreover, ANGPTL4 is also significantly upregulated in LUAD samples, which are all closely related to the development and a poor prognosis of LUAD (Liu et al., 2022b; Yang et al., 2022). Tetsuro Taki et al. demonstrated the biological significance of regulating the TGF-β signal in the cancer cell matrix through the correlation verification of CD109 and LTBP1, and they indicated that the expression level of CD109 plays an important role in promoting the proliferation and diffusion of LUAD cells (Taki et al., 2020). Furthermore, Lee et al. demonstrated that C109 expression is correlated with the invasiveness and metastasis of LUAD. They observed that CD109 expression is mechanistically mediated by binding to EGFR to regulate AKT/mTOR signaling (Lee et al., 2020). Therefore, in this study, the selected FAM-related genes may all be involved in the progression of LUAD and can serve as biomarkers for the clinical diagnosis and treatment of cancer. Therefore, the 12 key genes were used to construct a RiskScore model, laying a foundation for the survival prediction and further study of LUAD.

LUAD tissues had a lower score of the FAM-related pathway compared to para-cancer tissues. This also indicated that abnormal FAM was involved in the progression and prognosis of LUAD, which is consistent with the characterization results of FAM in LUAD by Wang et al. (2022a). As a new treatment method, immunotherapy has become an effective strategy to treat cancers (Riley et al., 2019). The level of immune cell infiltration in the tumor microenvironment has also been used as an important indicator for assessing lung cancer (Liu et al., 2021c). There are many studies investigating the effect of FAM on immunotherapy in various cancers (Bleve et al., 2020). The differential expression of FASN is closely correlated with immune cell infiltration, and patients with a low expression of FASN have active response to immune checkpoint inhibitor treatment (Xiong et al., 2022). The analysis and comparison of various tumor-related studies showed that the upregulated FASN gene expression and activity is negatively correlated with tumor immune infiltration. The methylation of the FASN promoter in DNA can be used to serve as a new biomarker for cancer (Zhang et al., 2022c). Hence, a better understanding of the correlation between FAM and the immunological signature of the tumor microenvironment could

facilitate the identification of new therapeutic targets for improving clinical cancer therapies.

However, this study also had certain limitations. The tumor and gene sample data collected were all from the database with a small number of samples, which demanded further *in vivo* or *in vitro* validation experiments to verify the predictive performance of the prognostic model. At the same time, the specific mechanism of abnormal FAM in LUAD was not clearly studied, and its interactions and regulatory mechanisms should be explored in depth.

5 Conclusion

In summary, this study determined 12 key genes (SLC2A1, PKP2, FAM83A, TCN1, MS4A1, CLIC6, UBE2S, RRM2, CDC45, IGF2BP1, ANGPTL4, and CD109) using six machine learning methods. A RiskScore model was constructed based on the 12 key genes mentioned previously. The model can accurately predict the survival of LUAD patients. We demonstrated that a specific model based on FAM could provide significant benefits for the precision treatment of LUAD and was effective in improving the prediction of patients' prognoses.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/ Supplementary Material.

Author contributions

DC: conceptualization and writing–original draft. YZ: data curation and writing–original draft. WZ: methodology and writing–original draft. JL: software and writing–original draft. YB: conceptualization, data curation, and writing–review and editing.

Funding

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Amiri, M., Yousefnia, S., Seyed Forootan, F., Peymani, M., Ghaedi, K., and Nasr Esfahani, M. H. (2018). Diverse roles of fatty acid binding proteins (FABPs) in development and pathogenesis of cancers. *Gene* 676, 171–183. doi:10.1016/j.gene.2018.07.035

Barrett, T., Wilhite, S. E., Ledoux, P., Evangelista, C., Kim, I. F., Tomashevsky, M., et al. (2013). NCBI GEO: archive for functional genomics data sets--update. *Nucleic acids Res.* 41, D991–D995. (Database issue). doi:10.1093/nar/gks1193

Bhaoighill, M. N., and Dunlop, E. A. (2019). Mechanistic target of rapamycin inhibitors: successes and challenges as cancer therapeutics. *Cancer drug Resist. (Alhambra, Calif.)* 2 (4), 1069–1085. doi:10.20517/cdr.2019.87

Bleve, A., Durante, B., Sica, A., and Consonni, F. M. (2020). Lipid metabolism and cancer immunotherapy: immunosuppressive myeloid cells at the crossroad. *Int. J. Mol. Sci.* 21 (16), 5845. doi:10.3390/ijms21165845

Chang, L., Fang, S., Chen, Y., Yang, Z., Yuan, Y., Zhang, J., et al. (2019). Inhibition of FASN suppresses the malignant biological behavior of non-small cell lung cancer cells via deregulating glucose metabolism and AKT/ERK pathway. *Lipids health Dis.* 18 (1), 118. doi:10.1186/s12944-019-1058-8

Charoentong, P., Finotello, F., Angelova, M., Mayer, C., Efremova, M., Rieder, D., et al. (2017). Pan-cancer immunogenomic analyses reveal genotype-immunophenotype relationships and predictors of response to checkpoint blockade. *Cell. Rep.* 18 (1), 248–262. doi:10.1016/j.celrep.2016.12.019

Chen, J., Fu, Y., Hu, J., and He, J. (2022a). Hypoxia-related gene signature for predicting LUAD patients' prognosis and immune microenvironment. *Cytokine* 152, 155820. doi:10.1016/j.cyto.2022.155820

Chen, Z., Hu, Z., Sui, Q., Huang, Y., Zhao, M., Li, M., et al. (2022b). LncRNA FAM83A-AS1 facilitates tumor proliferation and the migration via the HIF-1α/glycolysis axis in lung adenocarcinoma. *Int. J. Biol. Sci.* 18 (2), 522–535. doi:10.7150/ijbs.67556

Choi, R. Y., Coyner, A. S., Kalpathy-Cramer, J., Chiang, M. F., and Campbell, J. P. (2020). Introduction to machine learning, neural networks, and deep learning. *Transl. Vis. Sci. Technol.* 9 (2), 14. doi:10.1167/tvst.9.2.14

Danilova, L., Ho, W. J., Zhu, Q., Vithayathil, T., De Jesus-Acosta, A., Azad, N. S., et al. (2019). Programmed cell death ligand-1 (PD-L1) and CD8 expression profiling identify an immunologic subtype of pancreatic ductal adenocarcinomas with favorable survival. *Cancer Immunol. Res.* 7 (6), 886–895. doi:10.1158/2326-6066.CIR-18-0822

Dash, T. K., Chakraborty, C., Mahapatra, S., and Panda, G. (2022). Gradient boosting machine and efficient combination of features for speech-based detection of COVID-19. *IEEE J. Biomed. health Inf.* 26 (11), 5364–5371. doi:10.1109/JBHI.2022.3197910

Denisenko, T. V., Budkevich, I. N., and Zhivotovsky, B. (2018). Cell death-based treatment of lung adenocarcinoma. *Cell. death Dis.* 9 (2), 117. doi:10.1038/s41419-017-0063-y

DeVore, G. R. (2017). Computing the Z Score and centiles for cross-sectional analysis: a practical approach. *J. ultrasound Med. official J. Am. Inst. Ultrasound Med.* 36 (3), 459–473. doi:10.7863/ultra.16.03025

Divine, G., Norton, H. J., Hunt, R., and Dienemann, J. (2013). Statistical grand rounds: a review of analysis and sample size calculation considerations for Wilcoxon tests. *Anesth. analgesia* 117 (3), 699–710. doi:10.1213/ANE.0b013e31827f53d7

Ganggayah, M. D., Taib, N. A., Har, Y. C., Lio, P., and Dhillon, S. K. (2019). Predicting factors for survival of breast cancer patients using machine learning techniques. *BMC Med. Inf. Decis. Mak.* 19 (1), 48. doi:10.1186/s12911-019-0801-4

Garcia, K. A., Costa, M. L., Lacunza, E., Martinez, M. E., Corsico, B., and Scaglia, N. (2022). Fatty acid binding protein 5 regulates lipogenesis and tumor growth in lung adenocarcinoma. *Life Sci.* 301, 120621. doi:10.1016/j.lfs.2022.120621

Hänzelmann, S., Castelo, R., and Guinney, J. (2013). GSVA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinforma.* 14, 7. doi:10.1186/1471-2105-14-7

Hao, X. L., Tian, Z., Han, F., Chen, J. P., Gao, L. Y., and Liu, J. Y. (2019). Plakophilin-2 accelerates cell proliferation and migration through activating EGFR signaling in lung adenocarcinoma. *Pathology, Res. Pract.* 215 (7), 152438. doi:10.1016/j.prp.2019.152438

Hutchinson, B. D., Shroff, G. S., Truong, M. T., and Ko, J. P. (2019). Spectrum of lung adenocarcinoma. *Seminars ultrasound, CT, MR* 40 (3), 255–264. doi:10.1053/j.sult.2018.11.009

Kanehisa, M., and Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. *Nucleic acids Res.* 28 (1), 27–30. doi:10.1093/nar/28.1.27

Kang, J., Choi, Y. J., Kim, I. K., Lee, H. S., Kim, H., Baik, S. H., et al. (2021). LASSO-based machine learning algorithm for prediction of lymph node metastasis in T1 colorectal cancer. *Cancer Res. Treat.* 53 (3), 773–783. doi:10.4143/crt.2020.974

Landi, L., and Cappuzzo, F. (2015). Experience with erlotinib in the treatment of non-small cell lung cancer. *Ther. Adv. Respir. Dis.* 9 (4), 146–163. doi:10.1177/1753465815588053

Lee, K. Y., Shueng, P. W., Chou, C. M., Lin, B. X., Lin, M. H., Kuo, D. Y., et al. (2020). Elevation of CD109 promotes metastasis and drug resistance in lung cancer via activation of EGFR-AKT-mTOR signaling. *Cancer Sci.* 111 (5), 1652–1662. doi:10.1111/cas.14373

Li, B., Yu, L., and Gao, L. (2022b). Cancer classification based on multiple dimensions: SNV patterns. *Comput. Biol. Med.* 151, 106270. doi:10.1016/j.compbiomed.2022.106270

Li, H., Guo, L., and Cai, Z. (2022c). TCN1 is a potential prognostic biomarker and correlates with immune infiltrates in lung adenocarcinoma. *World J. Surg. Oncol.* 20 (1), 83. doi:10.1186/s12957-022-02556-8

Li, Y., Xu, Y., Ma, Z., Ye, Y., Gao, L., and Sun, Y. (2022a). An XGBoost-based model for assessment of aortic stiffness from wrist photoplethysmogram. *Comput. methods programs Biomed.* 226, 107128. doi:10.1016/j.cmpb.2022.107128

Li, Z., and Zhang, H. (2016). Reprogramming of glucose, fatty acid and amino acid metabolism for cancer progression. *Cell. Mol. life Sci. CMLS* 73 (2), 377–392. doi:10.1007/s00018-015-2070-4

Liang, C., Wang, X., Zhang, Z., Xiao, F., Feng, H., Ma, Q., et al. (2020). ACOT11 promotes cell proliferation, migration and invasion in lung adenocarcinoma. *Transl. lung cancer Res.* 9 (5), 1885–1903. doi:10.21037/tlcr-19-509

Lin, W., Chen, Y., Wu, B., Chen, Y., and Li, Z. (2021). Identification of the pyroptosis-related prognostic gene signature and the associated regulation axis in lung adenocarcinoma. *Cell. death Discov.* 7 (1), 161. doi:10.1038/s41420-021-00557-2

Liu, J., Gu, M., Xue, Y., Wang, Q., Ren, Y., and Huang, W. (2021c). Clinical significance of PD-L1 expression and CD8-positive tumor-infiltrating lymphocytes in patients with cavitary lung adenocarcinoma. *Oncologie* 23 (3), 439–452. doi:10.32604/oncologie.2021.017220

Liu, J., Li, Z., Cheang, I., Li, J., and Zhou, C. (2022b). RNA-binding protein IGF2BP1 associated with prognosis and immunotherapy response in lung adenocarcinoma. *Front. Genet.* 13, 777399. doi:10.3389/fgene.2022.777399

Liu, J., Sun, G., Pan, S., Qin, M., Ouyang, R., Li, Z., et al. (2020). The Cancer Genome Atlas (TCGA) based m(6)A methylation-related genes predict prognosis in hepatocellular carcinoma. *Bioengineered* 11 (1), 759–768. doi:10.1080/21655979.2020.1787764

Liu, T. T., Li, R., Huo, C., Li, J. P., Yao, J., Ji, X. L., et al. (2021a). Identification of CDK2-related immune forecast model and ceRNA in lung adenocarcinoma, a pan-cancer analysis. *Front. Cell. Dev. Biol.* 9, 682002. doi:10.3389/fcell.2021.682002

Liu, X., Li, J., Wang, Q., Bai, L., Xing, J., Hu, X., et al. (2022a). Analysis on heterogeneity of hepatocellular carcinoma immune cells and a molecular risk model by integration of scRNA-seq and bulk RNA-seq. *Front. Immunol.* 13, 1012303. doi:10.3389/fimmu.2022.1012303

Liu, Z., Zhang, W., Cheng, X., Wang, H., Bian, L., Wang, J., et al. (2021b). Overexpressed XRCC2 as an independent risk factor for poor prognosis in glioma patients. *Mol. Med. Camb. Mass.)* 27 (1), 52. doi:10.1186/s10020-021-00316-0

López-Ozuna, V. M., Kogan, L., Hachim, M. Y., Matanes, E., Hachim, I. Y., Mitric, C., et al. (2021). Identification of predictive biomarkers for lymph node involvement in obese women with endometrial cancer. *Front. Oncol.* 11, 695404. doi:10.3389/fonc.2021.695404

Lu, H., Wu, J., Liang, L., Wang, X., and Cai, H. (2022b). Identifying a novel defined pyroptosis-associated long noncoding RNA signature contributes to predicting prognosis and tumor microenvironment of bladder cancer. *Front. Immunol.* 13, 803355. doi:10.3389/fimmu.2022.803355

Lu, L., Wang, H., Fang, J., Zheng, J., Liu, B., Xia, L., et al. (2022a). Overexpression of OAS1 is correlated with poor prognosis in pancreatic cancer. *Front. Oncol.* 12, 944194. doi:10.3389/fonc.2022.944194

Lu, Y., Chen, X., Liu, F., Yu, H., Zhang, Y., Du, K., et al. (2022c). Systematic pan-cancer analysis identifies CDC45 as having an oncogenic role in human cancers. *Oncol. Rep.* 48 (4), 185. doi:10.3892/or.2022.8400

Ma, C., Luo, H., Cao, J., Gao, C., Fa, X., and Wang, G. (2020). Independent prognostic implications of RRM2 in lung adenocarcinoma. *J. Cancer* 11 (23), 7009–7022. doi:10.7150/jca.47895

Maan, M., Peters, J. M., Dutta, M., and Patterson, A. D. (2018). Lipid metabolism and lipophagy in cancer. *Biochem. biophysical Res. Commun.* 504 (3), 582–589. doi:10.1016/j.bbrc.2018.02.097

Mudd, T. W., Jr., Lu, C., Klement, J. D., and Liu, K. (2021). MS4A1 expression and function in T cells in the colorectal cancer tumor microenvironment. *Cell. Immunol.* 360, 104260. doi:10.1016/j.cellimm.2020.104260

Patz, E. F., Jr., Pinsky, P., Gatsonis, C., Sicks, J. D., Kramer, B. S., Tammemägi, M. C., et al. (2014). Overdiagnosis in low-dose computed tomography screening for lung cancer. *JAMA Intern. Med.* 174 (2), 269–274. doi:10.1001/jamainternmed.2013.12738

Pei, S., Liu, T., Ren, X., Li, W., Chen, C., and Xie, Z. (2021). Benchmarking variant callers in next-generation and third-generation sequencing analysis. *Briefings Bioinforma.* 22 (3), bbaa148. doi:10.1093/bib/bbaa148

Peng, Y., Liu, C., Li, M., Li, W., Zhang, M., Jiang, X., et al. (2021). Identification of a prognostic and therapeutic immune signature associated with hepatocellular carcinoma. *Cancer Cell. Int.* 21 (1), 98. doi:10.1186/s12935-021-01792-4

Polonio-Alcalá, E., Porta, R., Ruiz-Martínez, S., Vásquez-Dongo, C., Relat, J., Bosch-Barrera, J., et al. (2022). AZ12756122, a novel fatty acid synthase inhibitor, decreases resistance features in EGFR-TKI resistant EGFR-mutated NSCLC cell models. *Biomed. Pharmacother. = Biomedecine Pharmacother.* 156, 113942. doi:10.1016/j.biopha.2022.113942

Relat, J., Blancafort, A., Oliveras, G., Cufí, S., Haro, D., Marrero, P. F., et al. (2012). Different fatty acid metabolism effects of (-)-epigallocatechin-3-gallate and C75 in adenocarcinoma lung cancer. *BMC cancer* 12, 280. doi:10.1186/1471-2407-12-280

Riley, R. S., June, C. H., Langer, R., and Mitchell, M. J. (2019). Delivery technologies for cancer immunotherapy. *Nat. Rev. Drug Discov.* 18 (3), 175–196. doi:10.1038/s41573-018-0006-z

Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., et al. (2015). Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic acids Res.* 43 (7), e47. doi:10.1093/nar/gkv007

Santos, C. R., and Schulze, A. (2012). Lipid metabolism in cancer. *FEBS J.* 279 (15), 2610–2623. doi:10.1111/j.1742-4658.2012.08644.x

Shen, J., Huang, J., Huang, Y., Chen, Y., Li, J., Luo, P., et al. (2022). Anlotinib suppresses lung adenocarcinoma growth via inhibiting FASN-mediated lipid metabolism. *Ann. Transl. Med.* 10 (24), 1337. doi:10.21037/atm-22-5438

Shi, J., Chen, Y., Peng, C., Kuang, L., Zhang, Z., Li, Y., et al. (2022). Advances in targeted therapy against driver mutations and epigenetic alterations in non-small cell lung cancer. *Oncologie* 24 (4), 613–648. doi:10.32604/oncologie.2022.027545

Skalniak, L., Koj, A., and Jura, J. (2013). Proteasome inhibitor MG-132 induces MCPIP1 expression. *FEBS J.* 280 (11), 2665–2674. doi:10.1111/febs.12264

Skřičková, J., Kadlec, B., Venclíček, O., and Merta, Z. (2018). Lung cancer. *Cas. Lek. ceskych* 157 (5), 226–236.

Streeb, D., Metz, Y., Schlegel, U., Schneider, B., El-Assady, M., Neth, H., et al. (2022). Task-based visual interactive modeling: decision trees and rule-based classifiers. *IEEE Trans. Vis. Comput. Graph.* 28 (9), 3307–3323. doi:10.1109/TVCG.2020.3045560

Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., et al. (2021). Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA a cancer J. Clin.* 71 (3), 209–249. doi:10.3322/caac.21660

Taki, T., Shiraki, Y., Enomoto, A., Weng, L., Chen, C., Asai, N., et al. (2020). CD109 regulates *in vivo* tumor invasion in lung adenocarcinoma through TGF-β signaling. *Cancer Sci.* 111 (12), 4616–4628. doi:10.1111/cas.14673

Utkin, L. V., and Konstantinov, A. V. (2022). Attention-based random forest and contamination model. *Neural Netw. official J. Int. Neural Netw. Soc.* 154, 346–359. doi:10.1016/j.neunet.2022.07.029

Van Calster, B., McLernon, D. J., van Smeden, M., Wynants, L., Steyerberg, E. W., and Topic Group 'Evaluating diagnostic tests and prediction models' of the STRATOS initiative (2019). Calibration: the Achilles heel of predictive analytics. *BMC Med.* 17 (1), 230. doi:10.1186/s12916-019-1466-7

Van Calster, B., Wynants, L., Verbeek, J. F. M., Verbakel, J. Y., Christodoulou, E., Vickers, A. J., et al. (2018). Reporting and interpreting decision curve analysis: a guide for investigators. *Eur. Urol.* 74 (6), 796–804. doi:10.1016/j.eururo.2018.08.038

van de Vijver, M. J., He, Y. D., van't Veer, L. J., Dai, H., Hart, A. A., Voskuil, D. W., et al. (2002). A gene-expression signature as a predictor of survival in breast cancer. *N. Engl. J. Med.* 347 (25), 1999–2009. doi:10.1056/NEJMoa021967

Wang, D., Xiao, F., Feng, Z., Li, M., Kong, L., Huang, L., et al. (2020). Sunitinib facilitates metastatic breast cancer spreading by inducing endothelial cell senescence. *Breast cancer Res. BCR* 22 (1), 103. doi:10.1186/s13058-020-01346-y

Wang, H., Fang, J., Wang, Y., Li, S., Wang, Z., He, W., et al. (2022b). Gene editing in non-small cell lung cancer: current application and future perspective. *Oncologie* 24 (1), 65–83. doi:10.32604/oncologie.2022.021863

Wang, S., Chen, A., Zhu, W., Feng, D., Wei, J., Li, Q., et al. (2022a). Characterization of fatty acid metabolism in lung adenocarcinoma. *Front. Genet.* 13, 905508. doi:10.3389/fgene.2022.905508

Wu, L., Wang, H., Xia, Y., Xi, R., and Cnv-Bac, (2020). CNV-BAC: copy number variation detection in bacterial circular Genome. *Bioinforma. Oxf. Engl.* 36 (12), 3890–3891. doi:10.1093/bioinformatics/btaa208

Xiong, Q., Feng, D., Wang, Z., Ying, Y., Xu, C., Wei, Q., et al. (2022). Fatty acid synthase is the key regulator of fatty acid metabolism and is related to immunotherapy in bladder cancer. *Front. Immunol.* 13, 836939. doi:10.3389/fimmu.2022.836939

Yang, W. X., Gao, H. W., Cui, J. B., Zhang, A. A., Wang, F. F., Xie, J. Q., et al. (2023). Development and validation of a coagulation-related genes prognostic model for hepatocellular carcinoma. *BMC Bioinforma.* 24 (1), 89. doi:10.1186/s12859-023-05220-4

Yang, Y., Liu, Y., Gao, P., Liu, K., Zhao, K., Ying, R., et al. (2022). Prognostic significance of ANGPTL4 in lung adenocarcinoma: a meta-analysis based on integrated TCGA and GEO databases. *Evidence-based complementary Altern. Med. eCAM* 2022, 3444740. doi:10.1155/2022/3444740

Yang, Z., Wei, X., Pan, Y., Xu, J., Si, Y., Min, Z., et al. (2021). A new risk factor indicator for papillary thyroid cancer based on immune infiltration. *Cell. death Dis.* 12 (1), 51. doi:10.1038/s41419-020-03294-z

Yu, F. Y., Xu, Q., Wei, Q. Y., Mo, H. Y., Zhong, Q. H., Zhao, X. Y., et al. (2022). ACC2 is under-expressed in lung adenocarcinoma and predicts poor clinical outcomes. *J. cancer Res. Clin. Oncol.* 148 (11), 3145–3162. doi:10.1007/s00432-021-03910-1

Yu, X. H., Ren, X. H., Liang, X. H., and Tang, Y. L. (2018). Roles of fatty acid metabolism in tumourigenesis: beyond providing nutrition (Review). *Mol. Med. Rep.* 18 (6), 5307–5316. doi:10.3892/mmr.2018.9577

Zhang, M., Liu, Y., Yin, Y., Sun, Z., Wang, Y., Zhang, Z., et al. (2022b). UBE2S promotes the development of ovarian cancer by promoting PI3K/AKT/mTOR signaling pathway to regulate cell cycle and apoptosis. *Mol. Med. Camb. Mass.)* 28 (1), 62. doi:10.1186/s10020-022-00489-2

Zhang, M. Y., Huo, C., Liu, J. Y., Shi, Z. E., Zhang, W. D., Qu, J. J., et al. (2021b). Identification of a five autophagy subtype-related gene expression pattern for improving the prognosis of lung adenocarcinoma. *Front. Cell. Dev. Biol.* 9, 756911. doi:10.3389/fcell.2021.756911

Zhang, M., Yu, L., Sun, Y., Hao, L., Bai, J., Yuan, X., et al. (2022c). Comprehensive analysis of FASN in tumor immune infiltration and prognostic value for immunotherapy and promoter DNA methylation. *Int. J. Mol. Sci.* 23 (24), 15603. doi:10.3390/ijms232415603

Zhang, W., Ji, L., Wang, X., Zhu, S., Luo, J., Zhang, Y., et al. (2021a). Nomogram predicts risk and prognostic factors for bone metastasis of pancreatic cancer: a population-based analysis. *Front. Endocrinol.* 12, 752176. doi:10.3389/fendo.2021.752176

Zhang, Y. P., Wang, X., Jie, L. G., Qu, Y., Zhu, X. T., Wu, J., et al. (2022a). Osteoarticular involvement-associated biomarkers and pathways in psoriasis: the shared pathway with ankylosing spondylitis. *Front. Immunol.* 13, 836533. doi:10.3389/fimmu.2022.836533

Zheng, X., Ma, Y., Bai, Y., Huang, T., Lv, X., Deng, J., et al. (2022). Identification and validation of immunotherapy for four novel clusters of colorectal cancer based on the tumor microenvironment. *Front. Immunol.* 13, 984480. doi:10.3389/fimmu.2022.984480

Zhou, S. (2022). Sparse SVM for sufficient data reduction. *IEEE Trans. pattern analysis Mach. Intell.* 44 (9), 5560–5571. doi:10.1109/TPAMI.2021.3075339

Zhou, Z., Song, Q., Yang, Y., Wang, L., and Wu, Z. (2022). Comprehensive landscape of RRM2 with immune infiltration in pan-cancer. *Cancers* 14 (12), 2938. doi:10.3390/cancers14122938

Zhuang, W., Sun, H., Zhang, S., Zhou, Y., Weng, W., Wu, B., et al. (2021). An immunogenomic signature for molecular classification in hepatocellular carcinoma. *Mol. Ther. Nucleic acids* 25, 105–115. doi:10.1016/j.omtn.2021.06.024