Metabolic state uncovers prognosis insights of esophageal squamous cell carcinoma patients

Tingze Feng[1,2†], Pengfei Li[1†], Siyi Li[1,2†], Yuhan Wang[2], Jing Lv[2], Tian Xia[2], Hoy-Jong Lee[3], Hai-Long Piao[1,2*], Di Chen[2*] and Yegang Ma[1*]

Abstract

Background Metabolite-protein interactions (MPIs) are crucial regulators of cancer metabolism; however, their roles and coordination within the esophageal squamous cell carcinoma (ESCC) microenvironment remain largely unexplored. This study is the first to comprehensively map the metabolic landscape of the ESCC microenvironment by integrating an MPI network with multi-scale transcriptomics data.

Methods First, we characterized the metabolic states of cells in ESCC using single-cell transcriptome profiles of key metabolite-interacting proteins. Next, we determined the metabolic patterns of each ESCC patient based on the composition of different metabolic states within bulk samples. Finally, the ESCC samples were clustered into unique subtypes.

Results Sixteen ESCC metabolic states across 7 cell types were identified based on the re-analysis of single-cell RNA-sequencing data of 208,659 cells in 64 ESCC samples. Each of the 7 cell types within the tumor microenvironment exhibited distinct metabolic states, highlighting the high metabolic heterogeneity of ESCC. Based on differences in the compositions of the metabolic states, 4 ESCC subtypes were identified in two independent cohorts ($n = 79$ and 119), which were associated with significant variations in prognosis, clinical features, gene expression, and pathways. Notably, the inactivation of cellular detoxification processes may contribute to the poor prognosis of ESCC patients.

Conclusions Overall, we redefined robust ESCC prognostic subtypes and identified key MPI pathways that link metabolism to tumor heterogeneity. This study provides the first comprehensive mapping of the ESCC metabolic microenvironment, offering novel insights into ESCC metabolic diversity and its clinical applications.

Keywords Esophageal squamous cell carcinoma, Metabolite and protein interactions, Prognosis

†Tingze Feng, Pengfei Li and Siyi Li contributed equally to this work.

*Correspondence:
Hai-Long Piao
hpiao@dicp.ac.cn
Di Chen
di.chen@dicp.ac.cn

Yegang Ma
mayegang@cancerhosp-lncmu.com
[1]Department of Thoracic Surgery, Liaoning Cancer Hospital & Institute,
Cancer Hospital of China Medical University, Shenyang 110042, China
[2]Dalian Institute of Chemical Physics, Chinese Academy of Sciences,
Dalian 116023, China
[3]School of Pharmacy, Sungkyunkwan University, Suwon
16419, Republic of Korea

Introduction

Esophageal cancer (EC) is a common malignant tumor, ranking ninth in incidence and sixth in cancer-related mortality worldwide [1]. Esophageal squamous cell carcinoma (ESCC) is the predominant histological subtype of EC, accounting for approximately 86% of cases in China [2]. Many patients are diagnosed at an advanced stage due to the asymptomatic nature of early-stage ESCC [3]. Over the years, conventional therapeutic strategies have shown continuous improvement, with chemotherapy providing a systemic approach to target rapidly dividing cancer cells. For instance, thalidomide, once banned for its teratogenic effects, has been repurposed as an anti-cancer agent with immunomodulatory properties, demonstrating efficacy against hematologic malignancies and solid tumors [4]. Similarly, metal-based complexes, such as those derived from L-glutamic acid with copper (II) and ruthenium (III), exhibit strong DNA-binding affinities [5]. Additionally, advancements in polymeric nano-medicines have enabled targeted drug delivery, allowing for enhanced drug accumulation at tumor sites [6]. Despite these advancements, the mortality rate of ESCC remains alarmingly high, primarily due to significant intra-tumoral molecular and mutational heterogeneity [7].

Metabolic reprogramming, a hallmark of cancer [8], drives cancer cell growth, accelerated progression, and malignant proliferation within the tumor microenvironment (TME) [9]. Cancer cells gain a survival advantage by optimizing their metabolism to adapt to the TME [9, 10]. Core metabolic processes, such as glycolysis and fatty acid oxidation, meet the energy demands and support anabolic activities required for rapid proliferation [9]. Furthermore, novel therapeutic strategies targeting the metabolic and molecular vulnerabilities of the tumors have been proposed [11]. Recent research highlights the potential of metal-based drugs and amino acid derivatives as promising candidates for targeted cancer therapies. For instance, glutamic acid sulphonamides have demonstrated notable anticancer properties, while pyrazoline-based copper (II) and nickel (II) complexes have exhibited enhanced antifungal activity [12, 13]. The versatility of glutamic acid derivatives, in particular, supports their use in constructing multifunctional agents that can target cancer-specific pathways [14].

In the context of ESCC, several abnormal metabolic pathways have been reported to drive rapid cancer cell proliferation, including SREBF1-dependent fatty acid metabolism [15], cholesterol metabolism [16], and arginine metabolism [17]. However, metabolic abnormalities in cancer cells are not limited to isolated alterations in a few pathways but instead represent comprehensive disruptions of the entire metabolic network. These disruptions are mediated by specific metabolite-protein interactions (MPIs) that profoundly influence cancer cell survival [10, 18, 19]. Consequently, the metabolism landscape, which encompasses metabolites, proteins, and their interactions, varies significantly across cancer types [20]. Despite these insights, no systematic study has yet explored the reciprocal regulation between the metabolic alterations of ESCC cells and the TME, a mechanism recognized as critical for driving cancer progression [21]. Therefore, gaining a deeper understanding of the molecular mechanisms underlying ESCC is essential for developing more effective therapeutic strategies.

To further explore the metabolic characteristics in the ESCC microenvironment, we developed a workflow for metabolite state-based ESCC subtyping and mechanistic investigation. First, we identified the metabolic states of cells using the single-cell RNA-seq data (scRNA-seq), focusing on the expression of genes encoding critical metabolite-interacting proteins (CMIPs), which are highly connected nodes in the MPI network [22]. Next, we deconvoluted the proportions of these metabolic states in two bulk RNA-seq cohorts to characterize the metabolic composition of ESCC samples. Finally, we identified four prognostically distinct ESCC subtypes and analyzed their biological and clinical characteristics based on the relative proportions of metabolic states. Notably, the ESCC subtype associated with poor prognosis exhibited significant hypoxia, epithelial-mesenchymal transition (EMT) activity, and inactivation of cellular detoxification pathways.

Methods

The collection of ESCC datasets

Publicly available ESCC cohort datasets were systematically screened, verified, and matched with corresponding clinical annotations. Three ESCC cohorts comprising 257 samples were included in the analysis. The scRNA-seq cohort GSE160269 ($n = 64$, tumor samples: $n = 60$, normal samples: $n = 4$) [23] was utilized to identify distinct metabolic states of individual cells. In addition, TCGA-ESCC ($n = 79$) and GSE53624 ($n = 119$) [4] were employed to define distinct metabolic subgroups. The expression profiles of the TCGA esophageal cancer (ESCA) cohort ($n = 173$) were obtained through the data portal (https://xenabrowser.net/datapages/) [24], along with somatic mutation data and clinical data from tumor samples. From this cohort, 79 ESCC tumor samples with annotations for squamous cell neoplasms and labeled survival time were extracted. All the information regarding these cohorts has been summarized in Table 1, and the overall analysis workflow is illustrated in the flowchart (Fig. 1).

Data processing for scRNA-seq

The scRNA-seq analysis was conducted using the *Seurat* package (version 4.3.0) [25]. Initially, a series of data

Table 1 Information regarding the ESCC cohorts used in this study

Cohorts	PMIDs	Data type	Cli_info	Sur_info	Soma_info	Direction
GSE160269	34,489,433	ScRNA-seq	√	×	×	Identification for Mcs
TCGA-ESCC	NA	Bulk RNA-seq	√	√	√	Metabolic subtyping
GSE53624	24,522,499	Bulk RNA-seq	√	√	×	Validation for subtyping

PMIDs correspond to the publications of the cohorts used in this study. Cli_info annotated whether the cohorts contained clinically relevant information. Sur_info annotated whether the cohorts contained patients' survival information. Soma_info documented whether the cohorts included patients' somatic mutation information. Direction described the purpose of cohorts in this paper
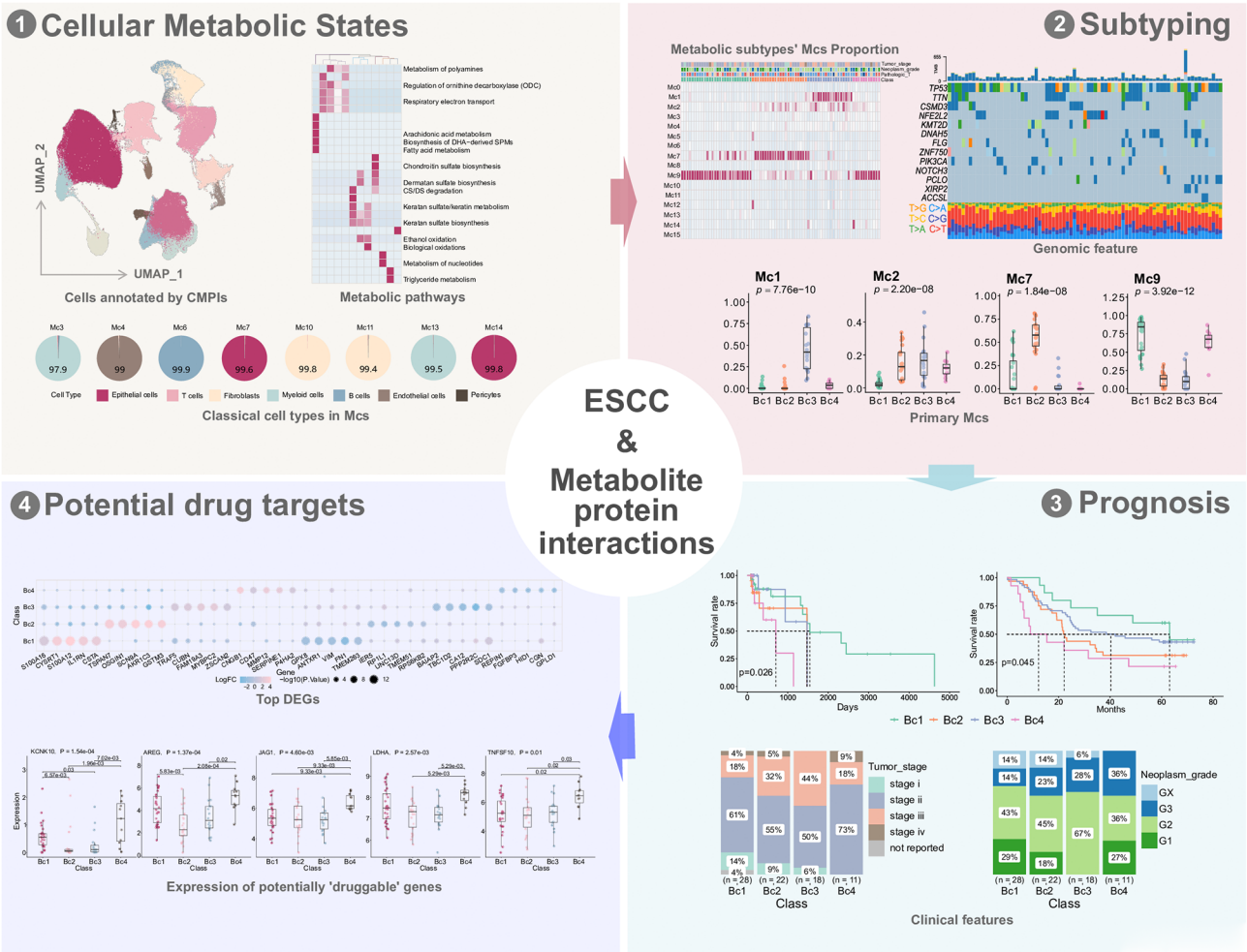
Fig. 1 Analytical workflow for annotating cellular metabolic states in ESCC and their association with patient prognosis

processing steps including data filtering, read normalization, construction of the KNN graph, identification of cell clusters, and Uniform Manifold Approximation and Projection (UMAP) dimensional reduction for both metabolic and typical cell clusters, were performed. The scRNA-seq data GSE160269 contains both CD45-positive and CD45-negative cells. After confirming the absence of batch effect between the samples and tissues (tumor and normal) in scRNA-seq data, CD45-positive and CD45-negative cells were merged for further analysis. For quality controls, the following criteria were applied: (1) cells with fewer than 500 unique feature counts were filtered out, (2) features expressed in fewer than three cells were excluded, and (3) all ribosome genes were removed from the dataset. The *SCTransform* function was used to normalize the raw unique molecular identifier counts. Deduction and clustering analyses were performed based on the top 2000 variable genes or genes encoding the CMIPs. Linear dimension reduction was executed using the *RunPCA* function, and the first 20 principal components (PCs) were selected as the optimal number of PCs, determined by generating an 'Elbow plot'. These 20 PCs were subsequently used to construct a KNN graph. The SLM algorithm was employed to cluster

the cells, with the resolution parameter set to 0.2. To further visualize cell clusters, the UMAP reduction was implemented by the *RunUMAP* function.

Cell type annotation
For typical cell annotation, the top 2000 variable features of scRNA-seq data were used to identify clusters. Typical cell types were annotated based on the expression of known marker genes [23]. The markers used for annotation included: *CD2, CD3D, CD3E,* and *CD3G* for T cell; *CD19, CD79A, MS4A1* for B cell; *EPCAM, SFN, KRT5,* and *KRT14* for epithelial cell; *CD68, LYZ, CD14, IL3RA, LAMP3, CLEC4C,* and *TPSAB1* for myeloid cell; *FN1, DCN, COL1A1, COL1A2, COL3A1* and *COL6A1* for fibroblast cell; *MCAM, RGS5,* and *ACTA2* for pericytes.

Identification of cellular metabolic states
CMIPs, which occupy hub positions in the MPI network, were identified as described by our previous study [22]. To determine the metabolic states of various cell types, the genes mapped to the CMIPs were utilized to find clusters.

Description of the metabolic patterns of ESCC patients based on bulk transcriptomics
Based on a study evaluating the performance of various deconvolution methods [26], the *DWLS* package (version 0.1.0) [27] was selected to deconvolute predicted metabolic cell fractions from TCGA-ESCC and GSE53624 datasets. All normal cells were excluded, and 20% of the cells in each metabolic state were extracted. The two bulk transcript profiling datasets were reversed and log-transformed prior to analysis. The deconvolution process was carried out following the *DWLS* reference manual. The signature matrix from single-cell data was constructed using the *buildSignatureMatrixMast* function. Subsequently, both the signature and bulk datasets were trimmed to include only the same differentially expressed genes using the *trimData* function. Finally, the metabolic cell fractions were estimated using the dampened weighted least squares method implemented by the *solveDampenedwls* function.

Metabolic pattern-based subtyping
Hierarchical clustering was performed using the R package *ConsensusClusterPlus* (version 1.16.0) based on each state's metabolic proportions of ESCC samples [28]. The process was repeated 10,000 times to ensure the stability of the classification. The matrix of metabolic cell proportions was normalized via *scale* function. Based on the consensus summary statistics, the samples were clustered into four groups (Fig S1).

Prediction of the four metabolic-related subtypes in another ESCC cohort
A linear discriminant analysis (LDA) model was trained to classify subtypes based on metabolic state proportions using TCGA-ESCC patient samples. The *Preprocess* and *predict.preProcess* functions were employed to preprocess the training data (TCGA-ESCC) and prediction data (GSE53624). A 10-fold cross-validation method was adopted. The model was trained and tested using the *caret* package (version 6.0–93).

Discerning subtype-relevant hub pathways
DEGs for each subgroup were identified using the R package *limma* (version 3.54.0) [29]. Gene set enrichment analysis (GSEA) based on the Reactome database was performed using *ReactomePA* (version 1.40.0) [30]. Additionally, hallmark-based pathway enrichment analysis (MSigDB 2023.1, https://www.gsea-msigdb.org/gsea/msigdb/human/collections.jsp#H) was performed using the *clusterProfiler* (version 4.9.0.002) [31, 32].

Gene set variation analysis
The pathway activity of REACTOME_BIOLOGICAL_OXIDATIONS or HALLMARK_EPITHELIAL_MESENCHYMAL_TRANSITION in each ESCC patient was estimated using the gene set variation analysis (GSVA) method implemented in the R package *GSVA* (version 1.52.3). The parameter 'method' was set to "ssgsea".

Statistical analysis
Statistical analyses and visualization results were conducted using R (version 4.2.0). The *ggplot2* package (version 3.4.2), *ggpubr* package (version 0.4.0), and *ggstatsplot* [33] package (version 0.9.5) were used for data visualization. Survival analysis was performed using the *survival* [34] (version 3.2–13) and the *survminer* (version 0.4.9) [35] packages. Detailed information, including data presentation, sample sizes, and statistical methods, is provided in relevant sections and figure legends.

Results
scRNA-seq profiles of ESCC revealed sixteen distinct cellular metabolic states
First, scRNA-seq profiles were annotated into seven typical cell types: epithelial cells, T cells, fibroblasts, myeloid cells, B cells, endothelial cells, and pericytes (Fig. 2A) based on clustering analysis. To investigate the metabolic states of the cells within the ESCC tumor microenvironment, the cells were clustered based on the expressional profiles of the CMIPs, which are the hub genes in the MPI network, as described in our previous study [22] (Table S1). This analysis identified 16 metabolic state clusters (Mcs) from Mc0 to Mc15 (Fig. 2B), indicating
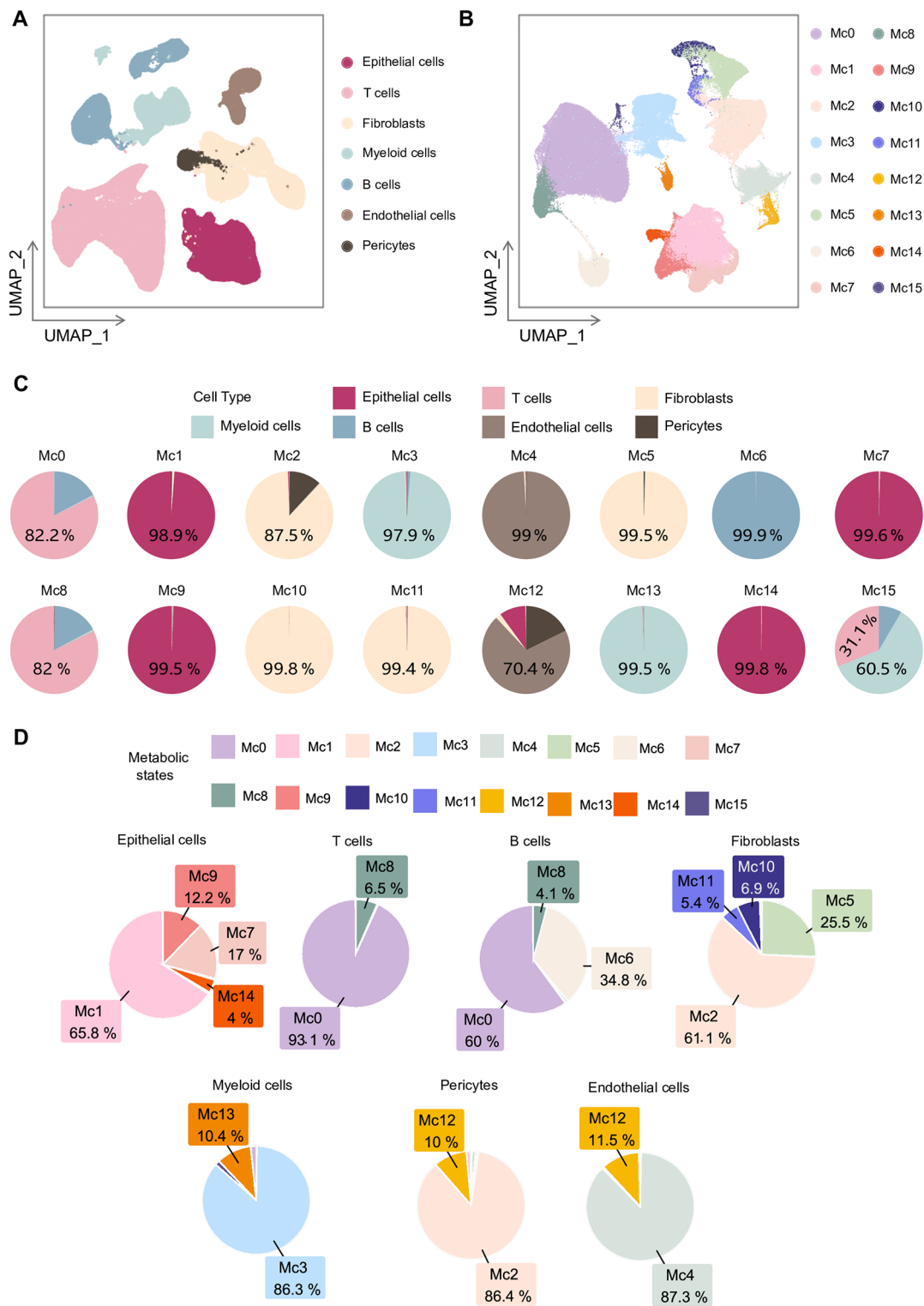
Fig. 2 The association between ESCC metabolic states and ESCC typical cell types. (**A**) UMAP plot of ESCC cells annotated based on canonical marker genes. (**B**) UMAP plot of ESCC cells annotated based on CMIPs. (**C**) Pie chart showing the proportions of the classical cell types within each metabolic state. (**D**) Pie chart showing the proportions of 16 metabolic states across seven classical cell types

potential metabolic heterogeneity within the ESCC tumor microenvironment.

To further elucidate the association between the seven typical cell types and these metabolic states, we calculated two proportions: the type-I proportions represent the proportions of typical cell types included in each metabolic state (Fig. 2C), while the type-II proportions represent the proportions of the metabolic states included within each typical cell type (Fig. 2D). The analysis showed that most metabolic states were dominated by a single cell type (e.g., one cell type accounted for more than 97% of the composition in 11 metabolic states). For example, B cells composed 99.9% of Mc6, while Mc3 and Mc13 were predominantly composed of myeloid cells (>97%), and Mc1, Mc7, Mc9, and Mc14 were composed almost entirely of epithelial cells (>98%) (Fig. 2C). However, there were still five metabolic states containing two or more types of cells (Fig. 2C). For instance, Mc2, Mc5, Mc10, and Mc11 were primarily composed of fibroblasts (87%), and Mc0 and Mc8 were predominantly composed of T cells (>82%) (Fig. 2C).

Furthermore, each typical cell type exhibited two or more metabolic states (Fig. 2D). Immune T and B cells, isolated from CD45$^+$ cells, shared two metabolic states, Mc0 and Mc8, suggesting that cells involved in the same biological process may exhibit similar metabolic characteristics (Fig. 2D). Additionally, 34.8% of B cells displayed the distinct metabolic state, Mc6. Both epithelial cells and fibroblasts were classified into four metabolic states (Fig. 2D). Notably, epithelial cells were characterized by four unique metabolic states, including Mc1 (65.8%), Mc7 (17%), Mc9 (12.2%), and Mc14 (4%), which were exclusively observed in epithelial cells (Fig. 2D). Similarly, fibroblasts were divided into Mc2 (61.1%), Mc5 (25.5%), Mc10 (6.9%), and Mc11 (5.4%) (Fig. 2D). Interestingly, 86.4% of pericytes exhibited the Mc2 metabolic state, suggesting that interacting cells may also share similar metabolic features.

This analysis revealed that different cell types not only shared metabolic states but also exhibited distinct, cell-type-specific metabolic states. These findings highlight the extensive intracellular and intercellular metabolic heterogeneity in the ESCC TME.

Each cell type of ESCC engaged in multiple metabolic pathways

We further examined the critical genes and pathways associated with 16 metabolic cellular states from two perspectives: typical cell types and metabolic states. These metabolic states displayed variations in several metabolism-related genes. For instance, the aldo/keto reductase superfamily genes *AKR1C1, AKR1C2, and AKR1C3*, which serve as marker genes for Mc7, were significantly enriched in this metabolic state. Similarly, other

metabolic states, such as Mc3, Mc8, and Mc11, showed elevated expression of specific gene families that function as marker genes (Fig. 3A, Table S2). Additionally, some metabolic states exhibited high expression levels of the classical cell-type markers. For example, Mc1, Mc7, Mc9, and Mc14 expressed epithelial cell marker genes prominently; Mc2, Mc5, Mc10, and Mc11 expressed fibroblast cell marker genes, and Mc3, Mc13, Mc15 expressed myeloid cell marker genes (Fig. 3B). These findings indicate that each cell type comprises multiple metabolic states, which explains why distinct marker genes were highly expressed across several metabolic states (Fig. 3A).

The pathway profiles of these identified metabolic states were also evaluated (see Methods). We found that several metabolic states were enriched in the common top pathways (Fig. 3C, D), suggesting that cells within these states likely perform similar functions. For instance, Mc1, Mc7, Mc8, Mc9, and Mc14 were enriched in the M phase and its downstream pathways. Furthermore, Mc1, Mc7, Mc9, and Mc14 were also enriched in the citric acid (TCA) cycle and respiratory electron transport pathways. In contrast, Mc2, Mc5, Mc10, and Mc11 were primarily associated with the extracellular matrix organization (Fig. 3C). The top five enriched pathways in Mc0, Mc3, and Mc15 were related to the immune system (Fig. 3C).

Focusing on the metabolism pathways, these metabolic states exhibited distinct profiles (Fig. 3D). For example, the Mc2, representing the metabolic state of some fibroblasts and pericytes, was highly enriched in chondroitin sulfate/dermatan sulfate metabolism and downstream pathways (Fig. 3D). Biological oxidation and ethanol oxidation pathways were activated in Mc5, another fibroblast's metabolic state. Arachidonic acid metabolism and its related pathways were significantly enriched in the Mc13(Fig. 3D).

In summary, distinct metabolic processes significantly contributed to the metabolic heterogeneity of the ESCC microenvironment.

MPI-related subtypes exhibited remarkable differences in prognosis

To determine whether metabolism heterogeneity is associated with ESCC prognosis, we performed deconvolution analyses of cellular metabolic states in bulk samples from the TCGA-ESCC cohort (Fig. 4A). Using the consensus clustering method based on the proportions of metabolic states in each sample, the TCGA-ESCC cohort was divided into four subtypes (Bulk cluster 1–4, Bc1-4), as indicated by changes in the cumulative distribution function and consensus matrix (Fig. S1). The analysis revealed that three metabolic states dominated by epithelial cells (Mc1, Mc7, Mc9) and one metabolic state dominated by fibroblasts (Mc2) contributed the majority for these subtypes (Fig. 4B, Fig. S2, Table S3). Specifically,
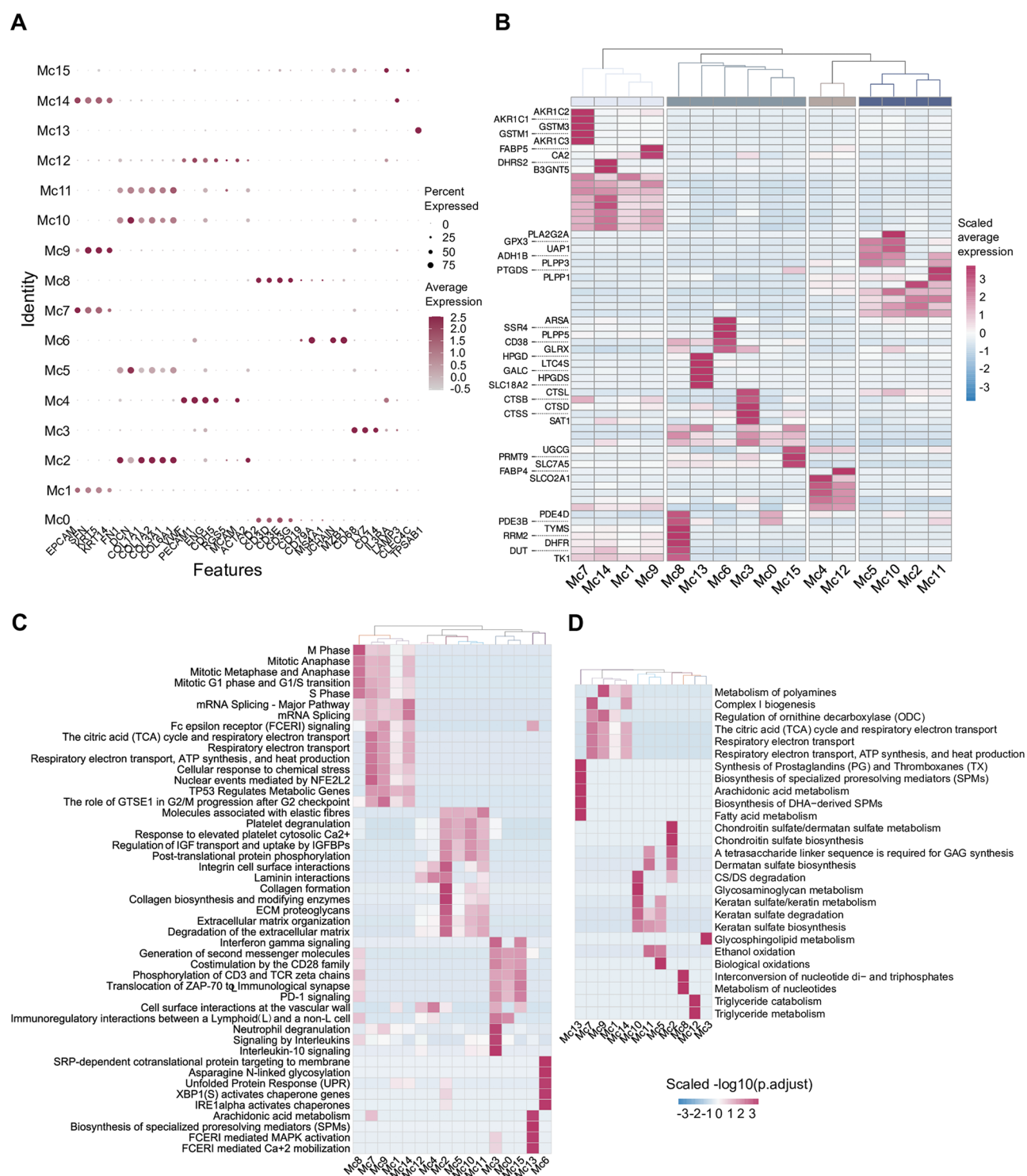
Fig. 3 Top expressed genes and pathways in each metabolic state. (**A**) Dot plot showing expression of canonical marker genes for 16 metabolic states. (**B**) Heatmap of top metabolic gene expressions for each metabolic state. Differentially expressed genes were identified using the *FindAllMarkers* function of the *Seurat* package. Only the top 5 differentially expressed metabolic genes were included. (**C-D**) The top five enriched pathways (**C**) and metabolic pathways (**D**) for each metabolic state. Adjusted p-values (p.adjust) were calculated using the *compareCluster* function of the *clusterProfiler* package, applying the Benjamini-Hochberg method. Over-representation analysis (ORA) of differentially expressed genes for each metabolic state was performed using genes with avg_log2FC > 0.8. The top 5 pathways for each metabolic state were ranked by p.adjust and clustered
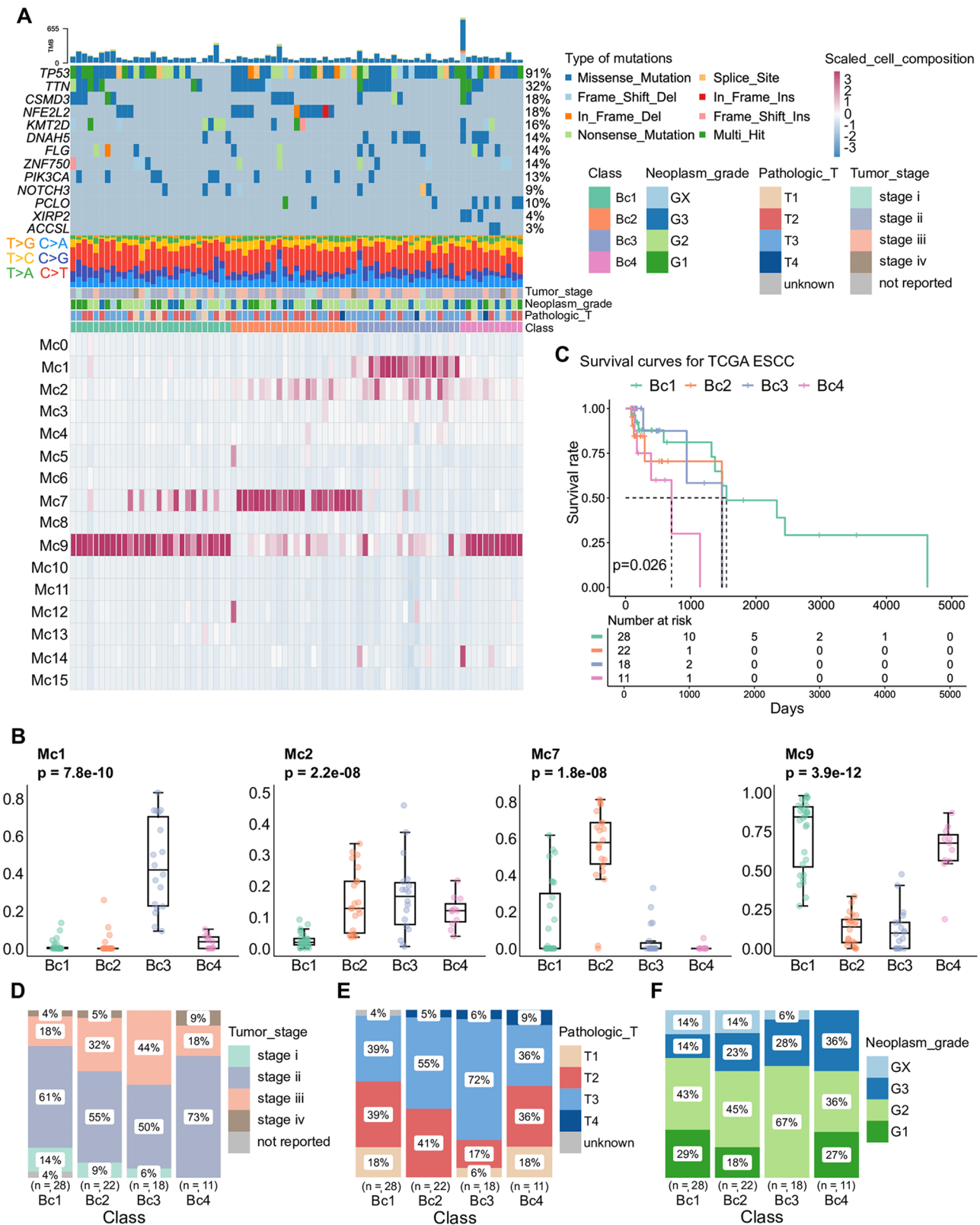
Fig. 4 (See legend on next page.)

(See figure on previous page.)

Fig. 4 Identification of four ESCC subtypes based on cellular metabolic state proportions. (**A**) Heatmap showing the genomic, clinical features, and metabolic state proportion differences among the four ESCC subtypes. The somatic mutation heatmap includes the top 10 mutated genes and subtype-specific mutated genes. (**B**) Box plot showing proportions of four central metabolic states across four bulk clusters (remaining significant Mc proportions were shown in Fig. S2). Statistical significance was assessed using the Kruskal–Wallis test (two-sided, unpaired), with P-values adjusted using the Holm method. In the boxplots, the central line represents the median, the bounds of boxes represent the first and third quartiles, and the upper and lower whiskers extend to the highest or the smallest value within the 1.5 interquartile range. (**C**) Kaplan-Meier (KM) plot showing the differential prognosis among the four ESCC subtypes (P-value calculated using log-rank test, two-sided). (**D-F**) Bar charts of different clinical features among patients in the four bulk clusters

Bc1 and Bc4 were dominated by Mc9, while Bc3 and Bc2 were predominantly influenced by Mc1 and Mc7, respectively. Notably, Mc2 was also relatively abundant in Bc2, Bc3, and Bc4 (Fig. 4B). Survival analysis showed that subtype Bc4, which primarily consisted of Mc2 and Mc9, had significantly worse prognosis outcomes compared to the other three subtypes (Fig. 4C). This suggests that the co-existence of the Mc2 and Mc9 metabolic states may be associated with poor prognosis.

Examining gene mutation profiles revealed that three genes (*PCLO*, *XIRP2*, and *ACCSL*) were significantly enriched in the subtype Bc4 (*PCLO*: $p = 0.0009$; *XIRP2*: $p = 0.002$; *ACCSL*: $p = 0.018$; fisher's exact test, two-sided). Missense mutations in *PCLO* have been implicated in promoting tumor aggressiveness in ESCC [36]. Similarly, mutations in *XIRP2* have been reported in ESCC, with a higher mutational frequency observed in older individuals [37]. Subtypes Bc2 and Bc3, which exhibited intermediate prognoses, showed alterations in *TP53* in all patients and had higher frequencies of missense mutations of *TTN* missense mutations compared to Bc1 and Bc4 (*TTN* mutation rates: 14% in Bc1, 27% in Bc2, 33% in Bc3, and 18% in Bc4). Notably, *NFEL2L2* mutations were most frequent in Bc2 (50%). In contrast, Bc1, which had the best prognosis, exhibited the lowest number of somatic mutations among subtypes. We further analyzed the TNM stages and metabolic subtypes in ESCC patients (Fig. 4D-F). Although patients in Bc4 had the worst prognosis, most were at tumor stage II, with pathologic T stages and neoplasm grades spanning T1 to T4 and G1 to G3 (Fig. 4D-F).

To validate the prognostic consistency of four subtypes, we assessed metabolic state proportions in an independent ESCC cohort, GSE53624 (Fig. 5A). Using an LDA-based metabolic state predictor constructed from the TCGA-ESCC cohort (see Methods), we predicted the metabolic subtypes for samples in the GSE53624 cohort. The distribution of metabolic state proportions across the four subtypes in GSE53624 closely resembled that of the TCGA-ESCC cohort (Fig. 5B, Fig. S3). Four metabolic states (Mc1, Mc2, Mc7, and Mc9) accounted for more than 60% of all cells. Consistent with the TCGA cohort, patients in Bc4 exhibited a poor prognosis compared to the other three subtypes, consistent with the TCGA-ESCC cohort (Fig. 5C, Table S3). Notably, Bc4 patients in

the GSE53624 cohort spanned all tumor grades, T stages, and N stages in the GSE53624.

These findings underscore that the changes in proportions of metabolic states are likely associated with the prognosis of ESCC patients.

Five significantly expressed genes of Bc4 in two cohorts were potentially 'druggable'

To identify potential therapeutic targets, we performed differential expression analyses comparing Bc4 with other subtypes and visualized the top five upregulated and downregulated protein-coding genes of two bulk cohorts (Fig. S4A, B, Table S2). We observed that genes differentially expressed were associated with disease progression, with distinct roles in tumor suppression or promotion. Subtypes Bc1 and Bc3 (with good prognosis) exhibited significantly expressed tumor suppressor genes, whereas DEGs in Bc2 and Bc4 (with poor prognosis) were associated with tumor cell proliferation. For instance, *S100A16*, the most significantly upregulated gene in Bc1, has been identified as a tumor suppressor in other squamous cell carcinomas. Its low expression was significantly correlated with reduced survival and poorer cancer differentiation [38]. Similarly, high levels of *S100A12* were associated with a favorable prognosis in SCC [39]. Additionally, low expression of *PKHD1L1* was linked to unfavorable overall survival in other cancers and was identified as a tumor suppressor gene [40]. In contrast, genes significantly upregulated in Bc2, such as *SCN9A* and *TXNRD1*, were found to promote tumor growth [41, 42]. In Bc4, *EGFR* and *CD47*, the most highly expressed genes, have been shown to promote cancer cell immune evasion [43].

Considering the poor prognosis of Bc4 patients, we identified five highly expressed genes that were drug-interaction targets in both cohorts, representing potential therapeutic candidates (Fig. S4C, D). All except *TNFSF10* [44] were newly identified as highly expressed in ESCC patients with poor prognosis. Notably, two of these drug-interaction genes were associated with metabolic processes. Lactate dehydrogenase A (LDHA), a critical enzyme in aerobic glycolysis [45], has been shown to promote tumor cell growth and cell migration in ESCC [46]. Several small-molecule inhibitors targeting *LDHA* have been developed as alternative anticancer therapeutics [45]. Another candidate, *KCNK10* also known as
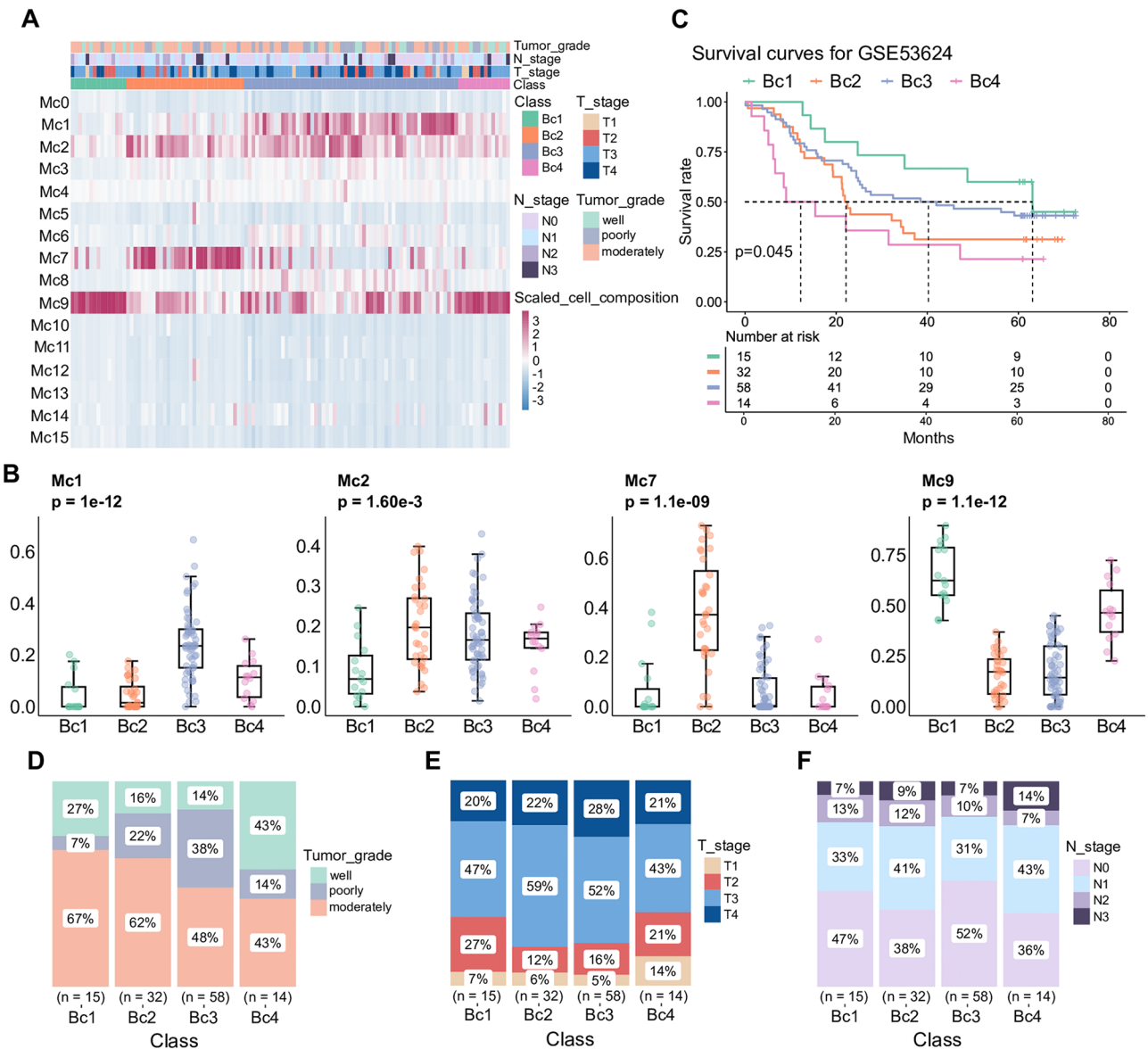
Fig. 5 Validation of four ESCC subtypes using GSE53624. (**A**) Heatmap showing clinical features and metabolic state proportion differences between the four ESCC subtypes. (**B**) Bar plot showing the proportions of four central metabolic states across four bulk clusters (remaining significant Mc proportions were shown in Fig. S3). Statistical significance was assessed using the Kruskal–Wallis test (two-sided, unpaired), with P-values adjusted using the Holm method. In the boxplots, the central line represents the median, the bounds of boxes represent the first and third quartiles, and the upper and lower whiskers extend to the highest or the smallest value within the 1.5 interquartile range. (**C**) KM plot showing the differential prognosis among the four ESCC subtypes. (P-value calculated using the log-rank test, two-sided). (**D-F**) Bar charts showing differences in clinical features among patients in the four bulk clusters

TREK2, is a two-pore domain potassium (K2P) channel regulated by several stimuli, such as fatty acids and stretch [47]. *KCNK10* also has been identified as a prognostic gene in other cancers [48].

Dysregulated cellular detoxification and EMT pathways in Bc4

To identify the discrepancies in biological processes between Bc4 and other subtypes, we performed pathway enrichment analysis on two cohorts. The results revealed that pathways related to cellular detoxification, such as glucuronidation (https://reactome.org/content/detail/RE ACT_6784) and glutathione conjugation (https://reacto me.org/content/detail/R-HSA-156590), were significantly downregulated in Bc4 across both two cohorts (Fig. 6A, Fig. S5, Fig. S6). In contrast, epithelial-mesenchymal transition (EMT) and hypoxia processes were markedly upregulated in Bc4 (Fig. 6B, Fig. S7, Fig. S8). To further illustrate these pathway alterations, we selected the biological oxidations gene set (including all genes associated

Fig. 6 (See legend on next page.)

(See figure on previous page.)

Fig. 6 Comparisons of significantly altered pathways among the four subtypes. (**A**) GSEA results for subtype Bc4 based on the Reactome database. NES: normalized enrichment score. Adjust p-value (p.adjust) was calculated using the *gsePathway* function from the *ReactomePA* package, applying the Benjamini-Hochberg method. (**B**) GSEA results for subtype Bc4 based on the HALLMARK set. Adjusted p-value (p.adjust) was computed using the *GSEA* function of the *clusterProfiler* package, applying the Benjamini-Hochberg method. (**C**) Box plot showing GSVA scores for biological oxidation across the four subtypes in two cohorts. Statistical significance was assessed using the Kruskal−Wallis (two-sided, unpaired), with P-values adjusted using the Holm method. (**D**) Box plot showing GSVA scores for EMT across the four subtypes in two cohorts. Statistical significance was assessed using the Kruskal−Wallis test (two-sided, unpaired), with P-values adjusted using the Holm method. In the boxplots, the central line represents the median, the bounds of boxes represent the first and third quartiles, and the upper and lower whiskers extend to the highest or the smallest value within the 1.5 interquartile range. (**E**) Dot plots showing marker genes for all epithelial cell metabolic states. (**F**) Dot plots showing marker genes of all fibroblast cell metabolic states. (**G**) ORA results for Mc9 (compared with other epithelial cell metabolic states). Adjusted p-value (p.adjust) was calculated using the *compareCluster* function of the *clusterProfiler* package, applying the Benjamini-Hochberg method. (**H**) ORA results of Mc2 (compared with other fibroblast cell metabolic states). Adjust p-value (p.adjust) was calculated using the *compareCluster* function of the *clusterProfiler* package applying the Benjamini-Hochberg method

with the cellular detoxification process) and the hypoxia gene set. Using GSVA scores, we compared these pathways across the two cohorts (Fig. 6C, D). The analysis showed significant differences in biological oxidation and EMT pathways between Bc4 and other subtypes (Fig. 6C, D). These findings suggest that the poor prognosis of Bc4 was associated with these dysregulated pathways.

Based on the metabolic cell proportion and prognosis of Bc4, we hypothesized that the inactivation of cellular detoxification pathways might be linked to the high proportions of Mc2 (fibroblast-dominated) and Mc9 (epithelial-dominated) states in Bc4. To test this hypothesis, we conducted differential expression analyses between Mc2 and other fibroblast metabolic states and between Mc9 and other epithelial cell metabolic states (Fig. 6E, F). Marker genes of cancer-associated fibroblasts (CAFs), such as *MMP1*, *ACTA2*, and *TAGLN* [23, 49–51], were predominantly expressed in Mc2. Notably, biological oxidation and its downstream pathways were downregulated in both Mc2 and Mc9, compared to their corresponding metabolic states in other cell types (Fig. 6G and H). These results confirmed our hypothesis that the simultaneous presence of high proportions of Mc2 and Mc9 in Bc4 leads to the inactivation of cellular detoxification processes, contributing to its poor prognosis.

Discussion

ESCC is the most common type of esophageal cancer in China and one of the leading causes of cancer-related mortality worldwide [2]. Metabolic reprogramming plays a critical role in cancer initiation, progression, and metastasis [52]. Previous studies have focused on metabolic reprogramming in ESCC, often examining the dysregulation of isolated metabolic pathways, such as dysregulated metabolic enzymes in glycolysis [53, 54] or fatty acid synthesis [55]. However, these studies have overlooked the high adaptability and complexity of metabolic reprogramming in the ESCC microenvironment. Specifically, they have failed to consider the intricate interactions between multiple metabolic pathways, which contribute to the overall metabolic landscape of the ESCC. This complexity can be reflected in the

activation of alternative metabolite-protein interactions (MPIs), which modulate the tumor's metabolic behavior and adapt it to the ever-changing microenvironment [21, 22]. Meanwhile, the scRNA-seq data have been widely applied to investigate the cell types in different tumors, however, the metabolic states of the cells were rarely studied. It limited our understanding of the metabolic reprogramming in the ESCC TME. Our study addresses this gap by integrative analyses of scRNA-seq and bulk RNA-seq data based on the MPI network. By leveraging the single-cell resolution of scRNA-seq data, we systematically characterized 16 cellular metabolic states and confirmed the high metabolic heterogeneity across different cell types within the ESCC TME (Fig. 2A). We provide a more global view of ESCC metabolism.

As an advanced technique, scRNA-seq enables the identification of specific cell types within the TME through unsupervised clustering [56]. Beyond defining cell types, scRNA-seq can also analyze cellular states from various perspectives, such as metabolism or proliferation. In this study, we analyzed the metabolic states of ESCC cells, offering a novel perspective for studying ESCC metabolism. We observed that each cell type exhibited multiple metabolic states, including shared metabolic states and cell-type-specific metabolic states. These patterns were observed in cells participating in similar physiological processes and interacting within the ESCC microenvironment. This finding underscores the extensive intracellular and intercellular metabolic heterogeneity in ESCC.

To further explore the relationship between ESCC prognosis and cellular metabolic states, we constructed a metabolism-related classification model. This model, validated by two bulk RNA-seq datasets, successfully deconvoluted the composition of metabolic states and identified novel metabolic-related subgroups. The four MPI-related subtypes exhibited significant differences in prognosis, suggesting that the metabolic states of ESCC patients directly influence outcomes. Interestingly, both Bc1 (associated with the best prognoses) and Bc4 (associated with the worst prognosis) had high proportions of Mc9. However, Bc4 exhibited a much higher proportion

of Mc2 compared to Bc1. Subtypes Bc2 and Bc3 also showed elevated Mc2 proportions, but their epithelial cells displayed other metabolic states (Fig. 4B). These findings suggested that the poor prognosis of ESCC patients may be associated with specific metabolic patterns of epithelial cells and fibroblasts.

To identify potential drug targets for delaying ESCC progression, we identified five significantly expressed 'druggable' genes in Bc4, the subtype with the worst prognosis. Gene set enrichment analysis revealed that cellular detoxification pathways were unexpectedly deactivated, while EMT hallmark pathways were significantly upregulated in Bc4. Differential expression analyses were performed to investigate the causes of these pathway dysregulations. Several cancer-associated fibroblast (CAF) marker genes, including *MMP1*, *ACTA2*, and *TAGLN*, were highly expressed in Mc2. Notably, Mc2 was composed of 87.5% fibroblasts and 12.5% pericytes. Since pericytes are a known source of CAFs [57], we hypothesize that Mc2 represents the metabolic state of CAFs. Our results highlight the critical role of specific metabolic patterns in determining ESCC patient outcomes. Specifically, the simultaneous inactivation of cellular detoxification in epithelial cells (Mc9) and fibroblasts (Mc2) with specific metabolic characteristics may contribute to the poor prognosis of ESCC patients.

MPIs are vital in enabling cell growth and survival in dynamic nutrient environments, contributing to tissue-specific metabolic flexibility [58]. However, few studies have identified specific MPI markers for particular cancers. For example, polyamine derivatives that bind CARTOR1 and Sestrins2 have demonstrated potential feedback regulation through the mTORC1 pathway, promoting polyamine synthesis in prostate cancer [59]. Similarly, pyruvate kinase M2 (PKM2) is highly expressed in human colorectal cancer cells [43], while LDHA-specific inhibitors have been shown to block aerobic glycolysis in cancer cells [58]. Consistent with these findings, hypoxia and glycolysis were significantly upregulated in the Bc4 subtype of the GEO cohort (Fig. S8). Hence, our study identified LDHA as a potential drug target for ESCC patients with poor prognosis.

Conclusion
This study systematically explored the metabolic characteristics of ESCC, identifying 16 distinct metabolic states that revealed extensive intracellular and intercellular metabolic heterogeneity within the ESCC microenvironment. By employing an LDA-based machine learning model, we proposed an effective ESCC subtyping strategy and validated significant prognostic differences among the four identified subtypes. Furthermore, we found that the inactivation of cellular detoxification pathways may be associated with poor prognosis in ESCC

patients. Notably, the high-degree genes from our MPI network and the comprehensive workflow presented here are broadly applicable, which can provide a framework to investigate metabolic characteristics and prognostics subtyping in other cancers.

Author contributions
T.Z., P.L., H-I.P., D.C., and Y.M. conceived and designed the project. T.Z., P.L., and S.L. conducted the bioinformatic analysis and generated data, and T.Z., H-I.P., D.C., and Y.M. wrote the manuscript. P.L., S.L., Y.W., J.L., T.X. H.L., and Y.M. contributed to data analysis and discussed the results. All authors read and approved the final manuscript.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare no conflict of interest.

References
1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global Cancer statistics 2020: GLOBOCAN estimates of incidence and Mortality Worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2021;71:209–49.
2. Chen R, Zheng R, Zhang S, Wang S, Sun K, Zeng H, et al. Patterns and trends in esophageal cancer incidence and mortality in China: an analysis based on cancer registry data. J Natl Cancer Cent. 2023;3:21–7.
3. Zhao Y-X, Zhao H-P, Zhao M-Y, Yu Y, Qi X, Wang J-H, et al. Latest insights into the global epidemiological features, screening, early diagnosis and prognosis prediction of esophageal squamous cell carcinoma. World J Gastroenterol. 2024;30:2638–56.

4. Ali I, Wani A, Saleem W, Haque K. Thalidomide: a banned drug resurged into Future Anticancer Drug. CDTH. 2012;7:13–23.
5. Ali I, Wani A, Saleem W, Wesselinova K, Syntheses D. DNA binding and anticancer profiles of L-Glutamic acid ligand and its copper(II) and ruthenium(III) complexes. Med Chem. 2013;9:11–21.
6. Ali I, Alsehli M, Scotti L, Tullius Scotti M, Tsai S-T, Yu R-S, et al. Progress in Polymeric Nano-Medicines for Theranostic Cancer Treatment. Polymers. 2020;12:598.
7. Dinh HQ, Pan F, Wang G, Huang Q-F, Olingy CE, Wu Z-Y, et al. Integrated single-cell transcriptome analysis reveals heterogeneity of esophageal squamous cell carcinoma microenvironment. Nat Commun. 2021;12:7335.
8. Hanahan D. Hallmarks of Cancer: New dimensions. Cancer Discov. 2022;12:31–46.
9. Dey P, Kimmelman AC, DePinho RA. Metabolic codependencies in the Tumor Microenvironment. Cancer Discov. 2021;11:1067–81.
10. Li Z, Sun C, Qin Z. Metabolic reprogramming of cancer-associated fibroblasts and its effect on cancer cell reprogramming. Theranostics. 2021;11:8322–36.
11. Ali I, Aboul-Enein H, Ghanem A. Enantioselective Toxic Carcinog CPA. 2005;1:109–25.
12. Ali I, Wani WA, Saleem K, Hsieh M-F. Anticancer metallodrugs of glutamic acid sulphonamides: in silico, DNA binding, hemolysis and anticancer studies. RSC Adv. 2014;4:29629–41.
13. Ali I, Wani WA, Khan A, Haque A, Ahmad A, Saleem K, et al. Synthesis and synergistic antifungal activities of a pyrazoline based ligand and its copper(II) and nickel(II) complexes with conventional antifungals. Microb Pathog. 2012;53:66–73.
14. Ali I, Wani WA, Haque A, Saleem K. Glutamic acid and its derivatives: candidates for Rational Design of Anticancer drugs. Future Med Chem. 2013;5:961–78.
15. Zhang G-C, Yu X-N, Guo H-Y, Sun J-L, Liu Z-Y, Zhu J-M, et al. PRP19 enhances esophageal squamous cell Carcinoma Progression by Reprogramming SREBF1-Dependent fatty acid metabolism. Cancer Res. 2023;83:521–37.
16. Tao M, Luo J, Gu T, Yu X, Song Z, Jun Y, et al. LPCAT1 reprogramming cholesterol metabolism promotes the progression of esophageal squamous cell carcinoma. Cell Death Dis. 2021;12:845.
17. Sun W, Kou H, Fang Y, Xu F, Xu Z, Wang X, et al. FOXO3a-regulated arginine metabolic plasticity adaptively promotes esophageal cancer proliferation and metastasis. Oncogene. 2024;43:216–23.
18. Faubert B, Solmonson A, DeBerardinis RJ. Metabolic reprogramming and cancer progression. Science. 2020;368:eaaw5473.
19. Wang W, Liu X, Chen H, Ling T, Xia T, Liu X, et al. Cholesterol as a functional metabolite cooperates with metadherin in cancer cells. Chin Chem Lett. 2020;31:1831–4.
20. Pavlova NN, Zhu J, Thompson CB. The hallmarks of cancer metabolism: still emerging. Cell Metabol. 2022;34:355–77.
21. Finley LWS. What is cancer metabolism? Cell. 2023;186:1670–88.
22. Chen D, Zhang Y, Wang W, Chen H, Ling T, Yang R, et al. Identification and characterization of Robust Hepatocellular Carcinoma Prognostic subtypes based on an integrative metabolite-protein Interaction Network. Adv Sci (Weinh). 2021;8:e2100311.
23. Zhang X, Peng L, Luo Y, Zhang S, Pu Y, Chen Y, et al. Dissecting esophageal squamous-cell carcinoma ecosystem by single-cell transcriptomic analysis. Nat Commun. 2021;12:5291.
24. Goldman MJ, Craft B, Hastie M, Repečka K, McDade F, Kamath A, et al. Visualizing and interpreting cancer genomics data via the Xena platform. Nat Biotechnol. 2020;38:675–8.
25. Hao Y, Hao S, Andersen-Nissen E, Mauck WM, Zheng S, Butler A, et al. Integrated analysis of multimodal single-cell data. Cell. 2021;184:3573–e358729.
26. Avila Cobos F, Alquicira-Hernandez J, Powell JE, Mestdagh P, De Preter K. Benchmarking of cell type deconvolution pipelines for transcriptomics data. Nat Commun. 2020;11:5650.
27. Tsoucas D, Dong R, Chen H, Zhu Q, Guo G, Yuan G-C. Accurate estimation of cell-type composition from gene expression data. Nat Commun. 2019;10:2975.
28. Wilkerson MD, Hayes DN. ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. Bioinformatics. 2010;26:1572–3.
29. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. Limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 2015;43:e47–47.
30. Yu G, He Q-Y. ReactomePA: an R/Bioconductor package for reactome pathway analysis and visualization. Mol BioSyst. 2016;12:477–9.
31. Wu T, Hu E, Xu S, Chen M, Guo P, Dai Z, et al. clusterProfiler 4.0: a universal enrichment tool for interpreting omics data. Innov. 2021;2:100141.
32. Xu S, Hu E, Cai Y, Xie Z, Luo X, Zhan L et al. Using clusterProfiler to characterize multiomics data. Nat Protoc [Internet]. 2024 [cited 2024 Jul 31]; Available from: https://www.nature.com/articles/s41596-024-01020-z
33. Patil I. Visualizations with statistical details: the ggstatsplot approach. JOSS. 2021;6:3167.
34. Therneau T. A Package for Survival Analysis in R. Available from: https://CRAN.R-project.org/package=survival
35. Alboukadel survminer. Drawing Survival Curves using ggplot2. 2020; Available from: https://CRAN.R-project.org/package=survminer
36. Sheng W, Li X, Li J, Mi Y, Li F. Evaluating prognostic value and relevant gene signatures of tumor microenvironment characterization in esophageal carcinoma. J Gastrointest Oncol. 2021;12:1228–40.
37. Li M, Zhang Z, Wang Q, Yi Y, Li B. Integrated cohort of esophageal squamous cell cancer reveals genomic features underlying clinical characteristics. Nat Commun. 2022;13:5268.
38. Basnet S, Vallenari EM, Maharjan U, Sharma S, Schreurs O, Sapkota D. An update on S100A16 in Human Cancer. Biomolecules. 2023;13:1070.
39. Hu Y, Han Y, He M, Zhang Y, Zou X. S100 proteins in head and neck squamous cell carcinoma (review). Oncol Lett. 2023;26:362.
40. Kang JY, Yang J, Lee H, Park S, Gil M, Kim KE. Systematic multiomic analysis of PKHD1L1 gene expression and its role as a Predicting Biomarker for Immune Cell Infiltration in skin cutaneous melanoma and Lung Adenocarcinoma. IJMS. 2023;25:359.
41. Bahcheli AT, Min H-K, Bayati M, Zhao H, Fortuna A, Dong W, et al. Pan-cancer ion transport signature reveals functional regulators of glioblastoma aggression. EMBO J. 2024;43:196–224.
42. Huang W, Liao Z, Zhang J, Zhang X, Zhang H, Liang H, et al. USF2-mediated upregulation of TXNRD1 contributes to hepatocellular carcinoma progression by activating Akt/mTOR signaling. Cell Death Dis. 2022;13:917.
43. Du L, Su Z, Wang S, Meng Y, Xiao F, Xu D, et al. EGFR-Induced and c-Src-mediated CD47 phosphorylation inhibits TRIM21-Dependent polyubiquitylation and degradation of CD47 to promote Tumor Immune Evasion. Adv Sci (Weinh). 2023;10:e2206380.
44. Zhang H, Qin G, Zhang C, Yang H, Liu J, Hu H, et al. TRAIL promotes epithelial-to-mesenchymal transition by inducing PD-L1 expression in esophageal squamous cell carcinomas. J Exp Clin Cancer Res. 2021;40:209.
45. Sharma D, Singh M, Rani R. Role of LDH in tumor glycolysis: regulation of LDHA by small molecules for cancer therapeutics. Sem Cancer Biol. 2022;87:184–95.
46. Li Q, Lin G, Zhang K, Liu X, Li Z, Bing X, et al. Hypoxia exposure induces lactylation of Axin1 protein to promote glycolysis of esophageal carcinoma cells. Biochem Pharmacol. 2024;226:116415.
47. Sorum B, Docter T, Panico V, Rietmeijer RA, Brohawn SG. Tension activation of mechanosensitive two-pore domain K + channels TRAAK, TREK-1, and TREK-2. Nat Commun. 2024;15:3142.
48. Cheng Y, Wang X, Qi P, Liu C, Wang S, Wan Q, et al. Tumor Microenvironmental competitive endogenous RNA network and Immune cells Act as Robust Prognostic Predictor of Acute myeloid leukemia. Front Oncol. 2021;11:584884.
49. Nurmik M, Ullmann P, Rodriguez F, Haan S, Letellier E. In search of definitions: Cancer-associated fibroblasts and their markers. Intl J Cancer. 2020;146:895–905.
50. Mathieson L, Koppensteiner L, Dorward DA, O'Connor RA, Akram AR. Cancer-associated fibroblasts expressing fibroblast activation protein and podoplanin in non-small cell lung cancer predict poor clinical outcome. Br J Cancer. 2024;130:1758–69.
51. Huang J, Tsang W-Y, Li Z-H, Guan X-Y. The origin, differentiation, and functions of Cancer-Associated fibroblasts in gastrointestinal Cancer. Cell Mol Gastroenterol Hepatol. 2023;16:503–11.
52. Martínez-Reyes I, Chandel NS. Cancer metabolism: looking forward. Nat Rev Cancer. 2021;21:669–80.
53. Liu J, Liu Z-X, Wu Q-N, Lu Y-X, Wong C-W, Miao L, et al. Long noncoding RNA AGPG regulates PFKFB3-mediated tumor glycolytic reprogramming. Nat Commun. 2020;11:1507.
54. Ma Y, Zhang X, Han X, Hao Z, Ji R, Li J, et al. PDK1 regulates the progression of esophageal squamous cell carcinoma through metabolic reprogramming. Mol Carcinog. 2023;62:866–81.
55. Cao S, Xue S, Li W, Hu G, Wu Z, Zheng J, et al. CircHIPK3 regulates fatty acid metabolism through miR-637/FASN axis to promote esophageal squamous cell carcinoma. Cell Death Discov. 2024;10:110.

Feng *et al. Journal of Translational Medicine*   (2025) 23:342

Page 15 of 15

56. Lei Y, Tang R, Xu J, Wang W, Zhang B, Liu J, et al. Applications of single-cell sequencing in cancer research: progress and perspectives. J Hematol Oncol. 2021;14:91.
57. Knipper K, Lyu S, Quaas A, Bruns C, Schmidt T. Cancer-Associated Fibroblast Heterogeneity and its influence on the Extracellular Matrix and the Tumor Microenvironment. IJMS. 2023;24:13482.
58. Hicks KG, Cluntun AA, Schubert HL, Hackett SR, Berg JA, Leonard PG, et al. Protein-metabolite interactomics of carbohydrate metabolism reveal regulation of lactate dehydrogenase. Science. 2023;379:996–1003.
59. Manning BD, Dibble CC. Growth Signaling Networks Orchestrate Cancer Metabolic Networks. Cold Spring Harb Perspect Med. 2024;14:a041543.

Publisher's note