




Identifying metabolic reprogramming phenotypes with glycolysis-lipid metabolism discoordination and intercellular communication for lung adenocarcinoma metastasis

Xin Li^{1,5}, Lefan Tang^{1,5}, Jiaying Deng¹, Xiuying Qi², Juxuan Zhang¹, Haitao Qi¹, Mengyue Li¹, Yixin Liu³, Wenyan Zhao¹, Yunyan Gu¹ , Lishuang Qi¹  ^{1✉} & Xia Li^{1,4}  [✉]

Tumor metastasis imposes metabolic requirements for escaping from primary tissues, producing vulnerability in treatment. This study aimed to explore the metabolic reprogramming relevant to lung adenocarcinoma (LUAD) metastasis and decode the underlying intercellular alterations. Using the gene expression profiles of 394 LUAD samples derived from The Cancer Genome Atlas (TCGA), we identified 11 metastasis-related metabolic genes involved in glycolysis and lipid metabolism, and defined three metabolic reprogramming phenotypes (MP-I, -II, and -III) using unsupervised clustering. MP-III with the highest glycolytic and lowest lipid metabolic levels exhibited the highest metastatic potency and poorest survival in TCGA and six independent cohorts totaling 1,235 samples. Genomic analyses showed that mutations in *TP53* and *KEAP1*, and deletions in *SETD2* and *PBRM1* might drive metabolic reprogramming in MP-III. Single-cell RNA-sequencing data from LUAD validated a metabolic evolutionary trajectory from normal to MP-II and MP-III, through MP-I. The further intercellular communications revealed that MP-III interacted uniquely with endothelial cells and fibroblasts in the ANGPTL pathway, and had stronger interactions with endothelial cells in the VEGF pathway. Herein, glycolysis-lipid dysregulation patterns suggested metabolic reprogramming phenotypes relevant to metastasis. Further insights into the oncogenic drivers and microenvironmental interactions would facilitate the treatment of LUAD metastasis in the future.

¹College of Bioinformatics Science and Technology, Harbin Medical University, Harbin 150086, China. ²Department of Anatomy, Harbin Medical University, Harbin 150081, China. ³Basic Medicine College, Harbin Medical University, Harbin 150086, China. ⁴Key Laboratory of High Throughput Omics Big Data for Cold Region's Major Diseases in Heilongjiang Province, Harbin, Heilongjiang 150081, China. ⁵These authors contributed equally: Xin Li, Lefan Tang. ✉email: qilishuang7@ems.hrbmu.edu.cn; lixia@hrbmu.edu.cn

Lung adenocarcinoma (LUAD) is the most common type of lung cancer, with a high mortality rate^{1,2}. Metastasis is the main cause of LUAD-related death, and the occult systemic spread of the disease in several patients cannot be detected through routine methods (such as pathological, clinical, and radiological evaluation)³. Therefore, a better understanding of the molecular mechanisms underlying LUAD metastasis is imperative for the identification of patients with occult metastases and for the development of customized treatment designs.

Metabolic reprogramming, conducted by cancer cells in response to stress, has been recognized as a requirement for the malignant progression of cancer, and its inhibition has been shown to reduce metastatic spread^{4,5}. Recent studies have demonstrated that the metabolic properties and propensities of tumors evolve during cancer progression, allowing tumor cells to acquire cell-autonomous properties associated with enhanced invasiveness or permitting them to alter the microenvironment to accelerate tumor metastasis^{6,7}, resulting in significant tumor metabolic heterogeneity. In tumor cells, the metabolic switch from oxidative phosphorylation to glycolysis (the Warburg effect) increases their cellular invasiveness and metastasis by promoting epithelial–mesenchymal transition and angiogenesis⁸. Other metabolic adaptations also contribute to the metastatic capabilities of tumor cells^{9,10}. Therefore, a more sophisticated view of cancer metastasis-related metabolic reprogramming is required.

The advent of transcriptomic technologies has provided a platform for the exploration of metabolic heterogeneity. A pan-cancer analysis has shown that tumor metabolic heterogeneity is associated with clinical outcomes¹¹. A study on pancreatic ductal adenocarcinoma reported a dysregulated glycolysis–cholesterol synthesis axis to be associated with patient outcome¹². However, cancer metabolism is context-specific⁶. Therefore, whether heterogeneity in distinct metabolic pathways can be used to stratify LUAD into clinically relevant phenotypes, has not yet been determined. Consequently, comprehensive interactions between metabolic phenotypes and the tumor microenvironment are not yet well understood. The emergence of single-cell transcriptomic sequencing (scRNA-seq) has accelerated the discovery of intercellular communications in context-specific diseases, making it possible to systematically elucidate the regulatory mechanisms of metabolic reprogramming in the tumor microenvironment.

This study aimed to identify the metastasis-related metabolic reprogramming phenotypes of LUAD. Genomic analyses were performed to unravel the underlying oncogenic events for the phenotypes. Further analyses of the scRNA-seq data of LUAD samples were used to validate the metabolic evolutionary trajectory from normal to metastatic phenotypes and reveal the intercellular communications.

Results

Identifying metastasis-related metabolic reprogramming phenotypes in LUAD. First, in The Cancer Genome Atlas (TCGA) dataset (Table 1), using Student's *t*-test with 5% false-discovery rate (FDR), we extracted 431 differentially expressed (DE) genes between the 185 metastatic LUAD samples with lymph node and/or distal metastasis and 209 non-metastatic LUAD samples (Fig. 1a), and found 280 and 151 DE genes to be upregulated and downregulated, respectively. Besides “Cell cycle” and “DNA replication” pathways, the upregulated DE genes also enriched in “Glycolysis/Gluconeogenesis” and “Carbon metabolism” pathways (Hypergeometric distribution model, FDR < 0.05, Fig. 1b). In addition, we found “HIF-1 signaling pathway”, associated with glycolysis, to be potentially enriched with the upregulated DE genes (Hypergeometric distribution model, $P < 0.05$, Fig. 1b). In contrast, biological pathways related to lipid metabolism,

including “Glycerophospholipid metabolism” and “Ether lipid metabolism”, were potentially enriched with the downregulated DE genes (hypergeometric distribution model, $P < 0.05$, Fig. 1c), indicating that lipid metabolism may play different roles from that of glycolysis in LUAD metastasis.

Therefore, we extracted the eight upregulated DE genes (*ALDOA*, *ENO1*, *GAPDH*, *GPI*, *LDHA*, *PGAM1*, *PGM2*, *TPI1*) from the “Glycolysis/Gluconeogenesis” pathway and three downregulated DE genes (*PLPP1*, *GPD1L*, *PLD3*) from the “Glycerophospholipid metabolism” and “Ether lipid metabolism” pathways (see Supplementary Table 1). The 11 DE genes were significantly positively correlated with each other within the same kind of pathway, while significantly negatively correlated with the genes in other kind pathways (Pearson correlation, FDR < 0.05, Supplementary Fig. 1). These results indicated that the imbalance of the glycolysis and lipid metabolism might participate in LUAD metastasis. Therefore, we performed unsupervised hierarchical clustering based on the mRNA expression of the 11 DE genes and found the samples to be clustered into three groups with different glycolytic and lipid metabolic patterns, defined as metabolic reprogramming phenotypes (MPs). The heatmap (Fig. 1d) showed that one MP exhibited lower mRNA expression of glycolytic genes accompanied with higher lipid metabolic genes; one MP displayed higher mRNA expression of glycolytic genes accompanied with lower lipid metabolic genes inversely, and the other MP showed intermediate mRNA expression of all metabolic genes. Whereafter, we calculated the MP-score of each group (see Methods), representing the imbalance of the two kinds of genes' expression, and defined the MP-I–III by scores from the lowest to the highest. Compared to the normal samples, the three phenotypes exhibited gradually increasing mRNA expression of glycolytic genes and decreasing expression of lipid metabolic genes (Fig. 1e).

We found MP-I–III to be significantly enriched in non-metastatic, lymph node metastatic, and distant metastatic samples, respectively (Fisher's exact test, $P < 0.0001$, Fig. 1f). Furthermore, we extracted the stage I treatment-naive patients and found that MP-III had significantly shorter OS than MP-I and MP-II (MP-III vs. MP-I: log-rank $P < 0.0001$; MP-III vs. MP-II: log-rank $P < 0.0001$, Fig. 1g), while there was no significantly different OS between MP-I and MP-II (log-rank $P = 0.4947$, Fig. 1g). Similarly, the percentages of distant metastatic samples were the highest (62.50%) in MP-III and lower in MP-II (27.03%) and in MP-I (13.04%) with a significant difference in the GSE11969¹³ dataset (Fisher's exact test, $P = 0.0169$, Supplementary Fig. 2a, b). Whereas the percentages of lymph node metastatic samples were not observed to be significantly different among the three phenotypes (Fisher's exact test, $P > 0.05$).

Depicting molecular characteristics and oncogenic events for metabolic phenotypes.

Differences in clinical and molecular characteristics (see Methods) across the metabolic phenotypes in the TCGA dataset were evaluated next, and displayed in Fig. 2 whose details are shown in Supplementary Table 2. The percentages of stage IV (15.49%) and stage III (30.99%) in MP-III was obviously higher than that in the other two MPs, the percentage of stage II (33.83%) was the highest in MP-II, whereas the percentage of stage I (58.15%) was the highest in MP-I (Fisher's exact test, $P < 0.0001$). Analogously, the percentages of magnoid subtype, squamoid subtype, and bronchioid subtype (a low invasion subtype¹⁴) were highest in MP-III (80.28%), MP-II (27.94%), and MP-I (66.31%), respectively (Fisher's exact test, $P < 0.0001$). The hypoxia score¹⁵, as well as stemness¹⁶ and proliferation scores¹⁷, increase progressively from MP-I to MP-III (Wilcoxon rank test, $P < 0.0001$). The immune score¹⁸ was the highest in MP-II, whereas the lowest was in MP-III (Wilcoxon rank test,

Table 1 The datasets of lung adenocarcinoma used in this study.

| Data source | Platform | Follow-up information | Sample count |
|-----------------|------------------------|-----------------------|--------------|
| TCGA | Illumina HiSeq 2000 | OS | 394 |
| GSE11969 | Agilent Homo 2.16 K | - | 90 |
| GSE31210 | Affymetrix Plus 2.0 | OS | 162 |
| GSE50081 | Affymetrix Plus 2.0 | OS | 92 |
| GSE13213 | Agilent 4x44K (G4112F) | OS | 79 |
| GSE42127 | Illumina WG-6 V3.0 | OS | 67 |
| GSE68465 | Affymetrix U133A | OS | 223 |
| GSE131907 | Illumina HiSeq 2500 | - | 11 |
| GSE123902 | Illumina HiSeq 2500 | - | 7 |
| CPTAC (mRNA) | Illumina HiSeq 4000 | - | 110 |
| CPTAC (protein) | Tandem mass tags | - | |

OS overall survival

$P < 0.0001$). In contrast, the tumor mutation burden (TMB) was observed to be highest in MP-III, while decreased to MP-II and MP-I (Wilcoxon rank test, $P < 0.0001$).

To determine the oncogenic events associated with the different metabolic phenotypes, we further investigated the differences in somatic mutations and copy number variations (CNVs) in the 11 metabolic genes and the 28 known LUAD driver genes (Supplementary Table 3) across the MPs in the TCGA dataset. We found the frequency of mutations in *TP53* and *KEAP1* and CNV deletion in a genomic region containing *SETD2* and *PBRM1* to be significantly different across the three phenotypes (Fisher's exact test, $FDR < 0.05$, Fig. 2). Of note, all the genetic events occurred more frequently in MP-III than in others.

Reconfirming the metabolic phenotypes of LUAD. The above result showed that a percentage of early-stage patients without any metastasis could be clustered into MP-III; thus, we inferred that these patients might have a high risk of metastasis, resulting in poor prognoses. Therefore, we first applied hierarchical clustering on stage I patients derived from five public datasets (Table 1) based on the mRNA expression of the 11 metabolic genes and confirmed the stage I patients also to be clustered into three phenotypes (Fig. 3a–e). Survival analysis (Fig. 3f–i) showed the stage I patients with MP-III to have significantly shorter OS than those with MP-I and/or MP-II in the GSE31210¹⁹ (MP-III vs. MP-I, log-rank $P < 0.0001$, HR = 3.00, 95% CI: 1.56–5.76; MP-III vs. MP-II: log-rank $P = 0.0015$, HR = 6.34, 95% CI: 1.72–23.44), GSE50081²⁰ (MP-III vs. MP-I, log-rank $P = 0.0174$, HR = 1.86, 95% CI: 1.08–3.20; MP-III vs. MP-II: log-rank $P = 0.0043$, HR = 3.85, 95% CI: 1.42–10.39), GSE13213²¹ (MP-III vs. MP-I, log-rank $P = 0.0412$, HR = 1.66, 95% CI: 1.00–2.76; MP-III vs. MP-II: log-rank $P = 0.4347$, HR = 1.58, 95% CI: 0.50–4.97), and GSE42127²² datasets (MP-III vs. MP-I, log-rank $P = 0.0243$, HR = 2.31, 95% CI = 1.02–5.26; MP-III vs. MP-II: log-rank $P = 0.0754$, HR = 2.64, 95% CI: 0.86–8.10). In the GSE68465²³ dataset, MP-III also exhibited shorter OS but without statistical significance (log-rank $P > 0.05$, Fig. 3j), which might be caused by the lack of one gene (*PGM2*) for clustering in the dataset. Additionally, there was no significant difference in OS between MP-I and MP-II in five independent datasets, but there was a shorter OS tendency of MP-II than that of MP-I in three datasets (GSE13213, GSE42127, and GSE68465). Especially, the GSE13213 dataset (Supplementary Fig. 2c, d) also showed that 30.77% stage I patients in MP-III developed distant metastasis after surgery, marginally significantly higher than that (8.11%) in MP-I (Fisher's exact test, $P = 0.0513$) and tentatively higher than that (22.22%) in MP-II (Fisher's exact test, $P = 0.66$).

Additionally, in a paired proteomic and transcriptomic dataset with 110 LUAD samples, derived from the Clinical Proteomic Tumor Analysis Consortium (CPTAC), we observed the samples were also clustered into the three MPs, based on the protein expression of the 11 metabolic genes (Fig. 4a). Moreover, we found that the mRNA expression of each metabolic gene was significantly positively correlated with its protein expression in the dataset (Pearson correlation, $FDR < 0.05$, Fig. 4b).

Displaying transcriptional trajectory of metabolic phenotypes at the single-cell level. From one scRNA-seq dataset (GSE131907²⁴), the 26,436 epithelial cells were cataloged into 14 clusters with 0.1 resolution that the normal cells (Cluster 1, 11, and 13), primary tumor cells (Cluster 4, 8, 9, and 10), and mBrain tumor cells (Cluster 0, 2, 3, 5, 6, 7, and 12) could optimally be separated, which was visualized by UMAP plot (Fig. 5a). Average mRNA expression of the eight glycolytic genes and three lipid metabolic genes for each cell is shown in Fig. 5b, and they were significantly higher and lower, respectively, in mBrain tumor cells than in primary tumor cells and normal cells. Therefore, we subjected the 11 metabolic genes to hierarchical clustering to determine the metabolic phenotypes of the epithelial tumor cells based on MP-score (Fig. 5c, d). The result confirmed the significantly higher percentage of mBrain tumor cells in MP-III than that in MP-I and MP-II (Fisher's exact test, both $P < 0.0001$, Fig. 5e).

We next constructed a metabolic trajectory (Fig. 5f) by ordering the epithelial cells according to the mRNA expression changes in the 11 metabolic genes (see Methods). Seven transcriptional states in the trajectory indicated distinct cellular metabolic fates. The percentages of each MP in the seven states are shown in Fig. 5g. Notably, State 1 and State 4 were both enriched with normal epithelial cells, while the cells in State 4 exhibited very low expression of all the 11 metabolic genes, which was not observed in the bulk dataset (TCGA), thus they were not analyzed in the further analyses. Here, the metabolic trajectory of the glycolysis-lipid imbalance appeared to begin principally from partial normal epithelial cells, which marked the beginning of State 1, evolved to MP-I tumor cells majorly at the end of State 1, and then formed a branched structure with two major cell fates (Cell fate 1 and 2). Tracing the metabolic trajectory of Cell fate 1 (ignoring State 4) revealed that MP-I tumor cells (State 2) evolved into MP-III tumor cells (State 3). Subsequently, following the trajectory of Cell fate 2, we found that along the MP-I tumor cell (State 5) trajectory, MP-II and MP-III tumor cells were mainly located in separate branches of State 6 and State 7 (Fig. 5g). The mRNA expression of glycolytic genes generally increased from pre-branch to State 6 (MP-II) and State 7 (MP-III), while a decrease in lipid metabolic gene expression was only observed from pre-branch to State 7

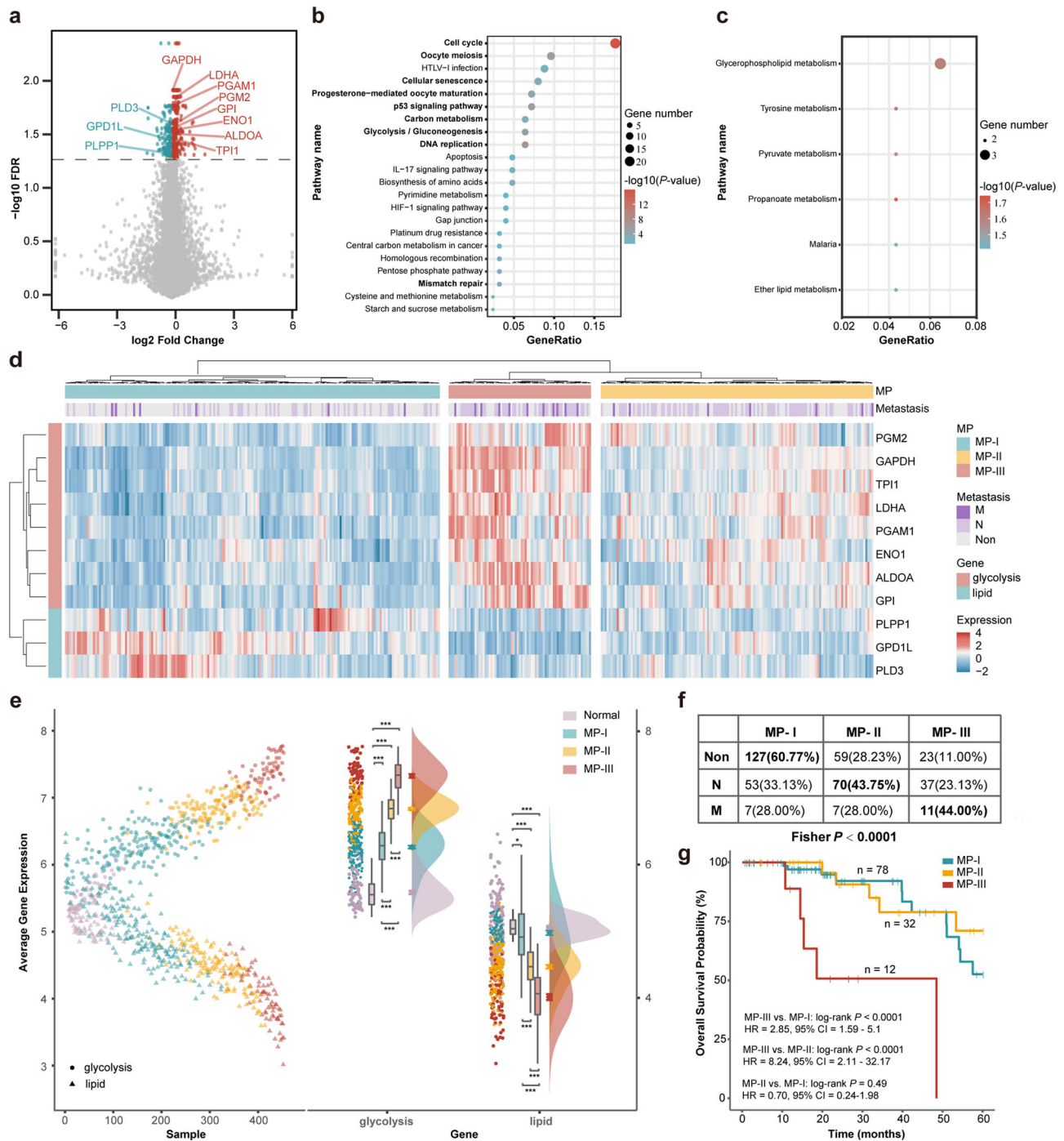


Fig. 1 Identification of metabolic reprogramming phenotypes (MPs) in the TCGA dataset. **a** Volcano plot of differentially expressed (DE) genes between metastatic and non-metastatic samples. Red and blue represented upregulated and downregulated DE genes, respectively. **b, c** Functional pathways enriched with upregulated (**b**) and downregulated (**c**) DE genes. **d** Hierarchical clustering heatmap for all the samples based on the mRNA expression (Z-score) of the 11 DE metabolic genes ($n = 394$ biologically independent samples). Here, the samples were clustered into the three MPs; the one with lower expression of glycolytic genes accompany by higher lipid metabolic genes were defined as MP-I, the one with higher expression of glycolytic genes accompanied with lower lipid metabolic genes was defined as MP-III, and the other one showing intermediate mRNA expression of all metabolic genes was defined as MP-II. **e** Average expression of glycolytic and lipid metabolic genes across the three MPs and normal samples. Boxplots extend from the 25th to 75th percentiles, the line indicates the median, and whiskers indicate the range. **f** Confusion matrix for the number of metastatic samples in the different MPs. **g** Kaplan-Meier curves of overall survival for 122 samples obtained from treatment-naive patients with stage I LUAD.

(Fig. 5h). A similar metabolic trajectory was observed from normal to mBrain (Supplementary Fig. 3).

Decoding the intercellular communications underlying metabolic phenotypes. To build a metabolic phenotype-specific

cell-cell communication atlas, we extracted 6372 tumor epithelial cells (MP-I, -II, and -III), 2373 stromal cells (fibroblasts and endothelial cells), and 35,506 immune cells (B lymphocytes, T lymphocytes, NK, myeloid, and MAST cells) from the primary tissues of patients with tLung. In total, we identified 41 significant

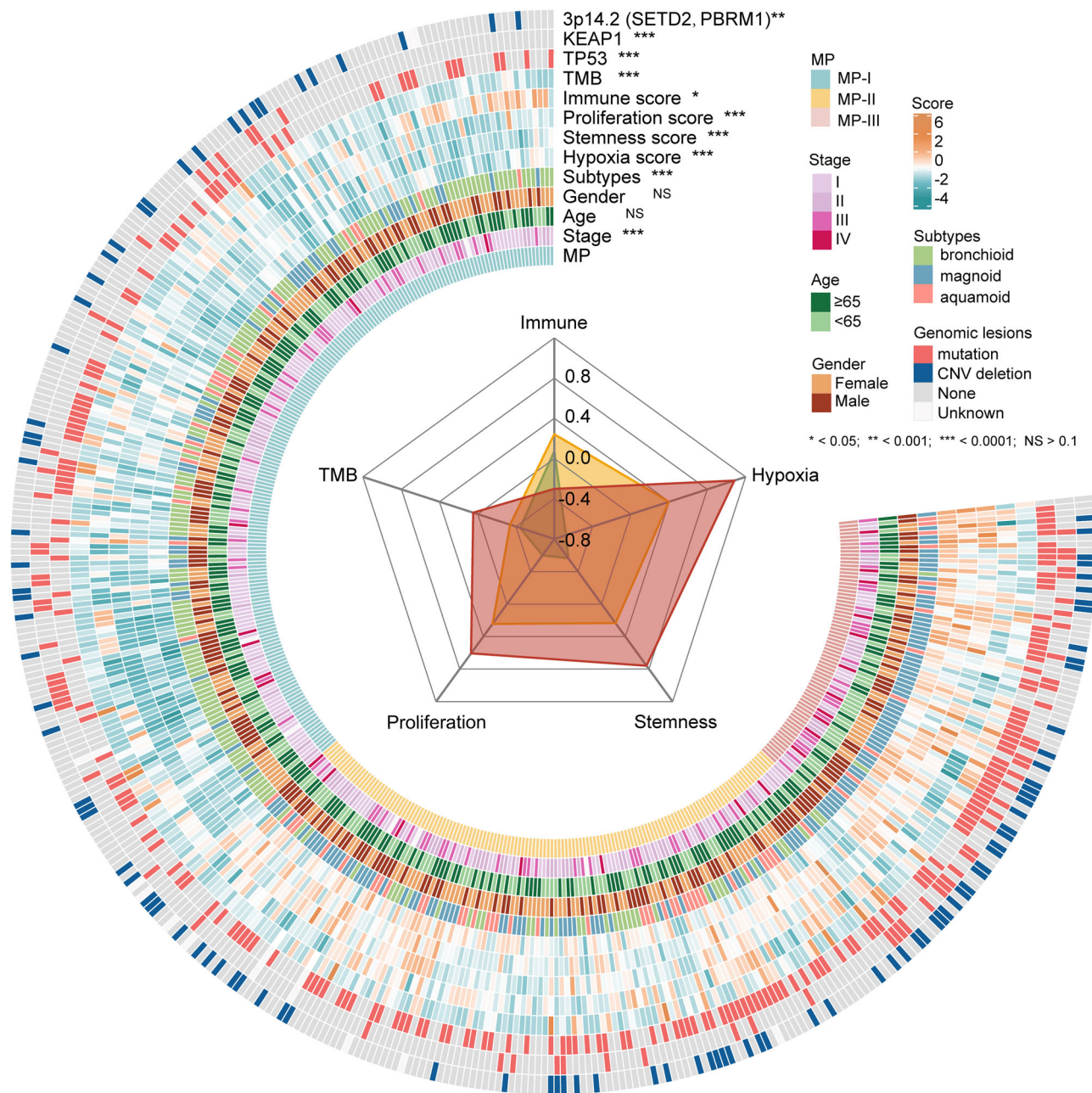


Fig. 2 Clinical and molecular characteristics of the MPs in the TCGA dataset. Circle heatmap of clinical information (stage, age, gender, transcriptomics subtypes, molecular scores, and TMB) and oncogenic events (*TP53*, *KEAP1*, *SETD2*, and *PBRM1*). The significance of differences (*P* value) among the three MPs, estimated by Fisher’s exact test, is shown after each characteristic. The radar chart (inner) displays four molecular scores and TMB of the three MPs. The detailed numbers of each characteristic in the MPs are presented in Supplementary Table 2.

ligand-receptor pairs between the 10 cell populations (Fig. 6a), and as expected the MP-III population showed the strongest outgoing signal, followed by MP-II, with MP-I weakest (Fig. 6b).

Next, significant ligand–receptor pairs associated with the three metabolic phenotypes were further categorized into 15 signaling pathways (Fig. 6c), including *ANGPTL*, *EGF*, *SPP1*, and *VEGF* pathways. Network centrality analysis of the inferred *ANGPTL* signaling network showed the MP-III population to be a unique source of *ANGPTL* ligands (*ANGPTL4*, *ANGPTL2*) that acted on endothelial cells and fibroblasts (Fig. 7a). The relative contributions of each ligand-receptor pair in the signaling pathway and the role of each cell population are shown in Fig. 7b, c. For the *VEGF* signaling pathway, all three MP populations were prominent sources of *VEGF* ligands acting on endothelial cells,

and the interaction intensity gradually increased from MP-I to MP-III (Fig. 7d–f). Furthermore, the three MP populations, especially MP-II, were found to be receivers in the *EGF* signaling pathway, possibly acted upon by themselves as well as by myeloid, MAST, NK, and endothelial cells (Fig. 7g–i). Notably, the MP-III population exhibited significant signal sending activities toward the MP-II population in the *EGF* signaling.

Collectively, we identified 103 ligand–receptor pairs interacting MP-III with other cell populations (Supplementary Data 1). Similarly, we identified 154 significant ligand–receptor pairs interacting MP-III with other cell populations in an independent scRNA-seq dataset (GSE123902²⁵). Thereinto, there were 62 overlapped ligand-receptor pairs, significantly higher than expected (hypergeometric distribution model, *P* < 0.0001, Supplementary

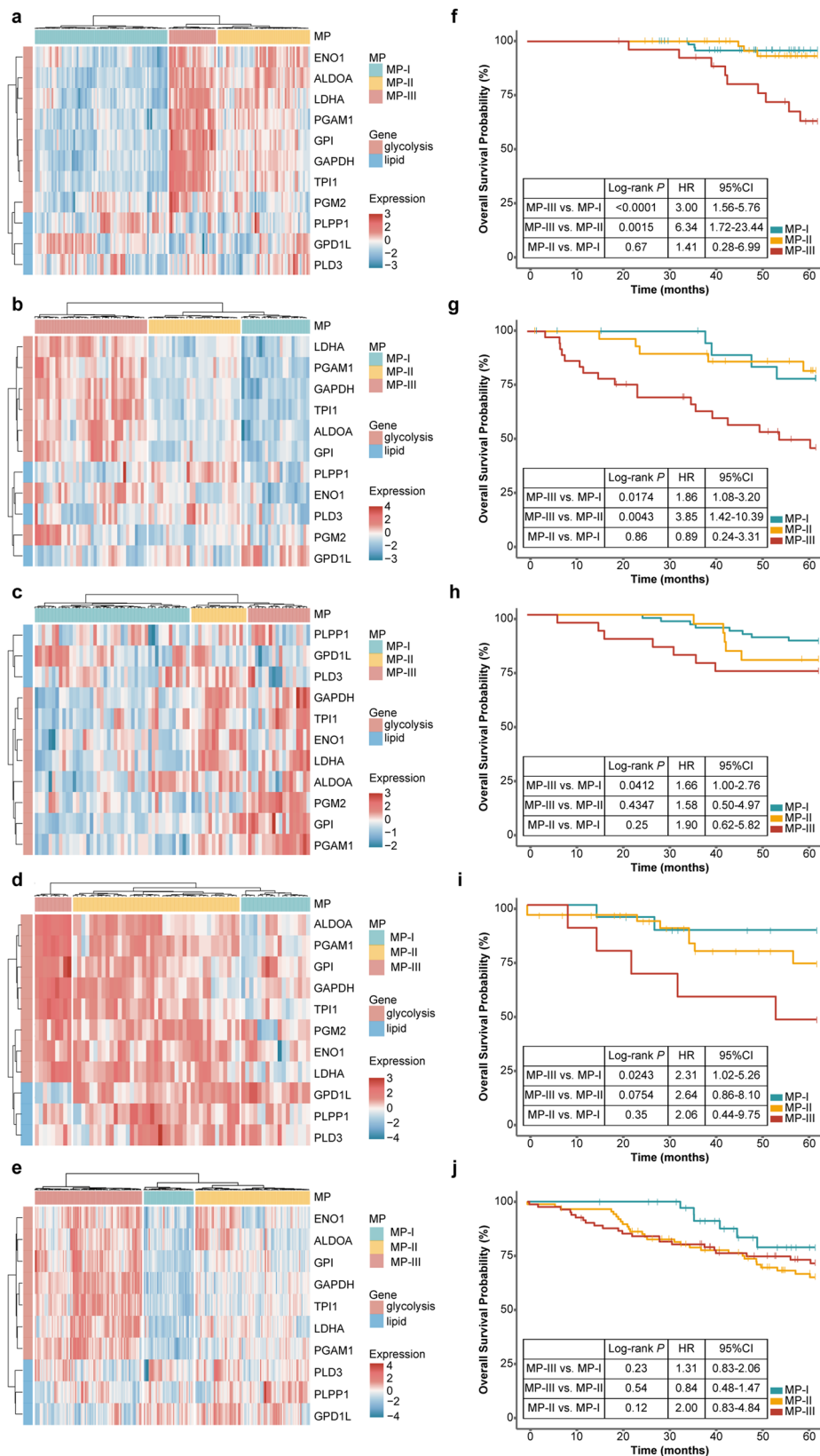


Fig. 3 Reconfirmation of the MPs and survival in multiple cohorts. **a–e** Hierarchical clustering heatmap for treatment-naïve patients with stage I LUAD based on the mRNA expression (Z-score) of the 11 metabolic genes in the GSE31210 ($n = 162$ biologically independent samples), GSE50081 ($n = 92$ biologically independent samples), GSE13213 ($n = 79$ biologically independent samples), GSE42127 ($n = 67$ biologically independent samples), and GSE68465 ($n = 223$ biologically independent samples) datasets. **f, j** Kaplan–Meier curves of overall survival for the five datasets.

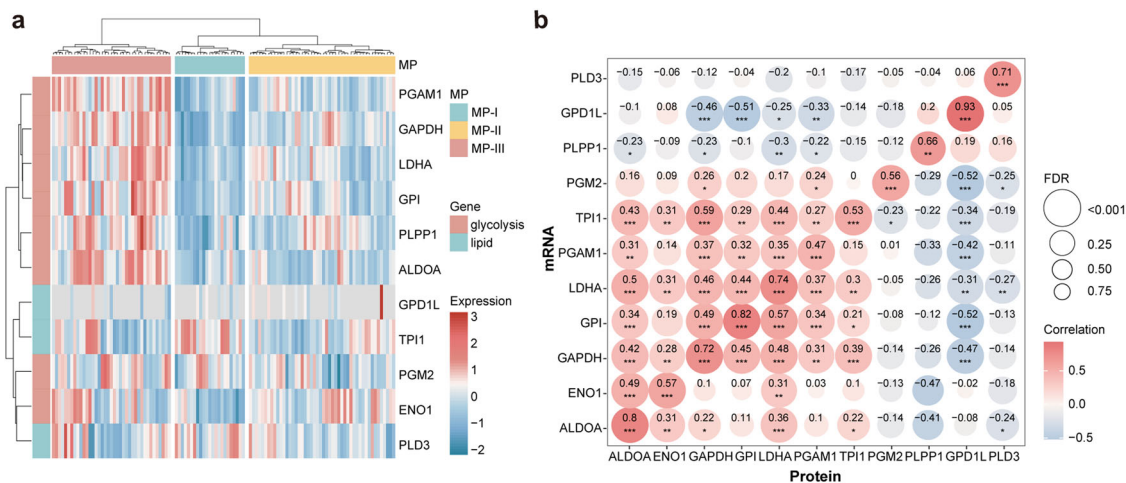


Fig. 4 Reconfirming of metabolic phenotypes in the proteomic dataset. **a** Hierarchical clustering heatmap based on the protein expression (Z-score) of the 11 metabolic genes from CPTAC. **b** Correlation matrix for the mRNA and protein expression of the 11 metabolic genes.

Fig. 4). Notably, mRNA expression of nine ligands expressed on MP-III, including *ANGPTL4*, *VEGFA*, etc., was significantly higher than that on MP-I in the GSE131907, which were all validated in another scRNA-seq dataset (Wilcoxon rank test, FDR < 0.05, Supplementary Table 4). Furthermore, in the TCGA dataset, we found the mRNA expression of all the nine ligands to be significantly correlated with the mRNA expression of 11 metabolic genes, which could be regulated by *TP53* and *KEAP1* mutations, and *SETD2* and *PBRM1* deletions (Pearson correlation, FDR < 0.05). A comprehensive network was constructed to link the intracellular oncogenic and metabolic events of MP-III with intercellular communications (Fig. 8).

Discussion

Metabolic reprogramming is a hallmark of tumor cells and has been recently regarded as a key contributor in tumor progression²⁶. In this study, we found that upregulated and downregulated DE genes in metastatic samples were enriched in glycolysis and lipid metabolism pathways, respectively, thereby indicating that the imbalance of glycolysis and lipid metabolism may promote LUAD metastasis. Glycolysis is known to facilitate tumor invasion and metastasis²⁷. Among the eight upregulated glycolytic genes, *ALDOA*, *GAPDH*, *LDHA*, and *PGAM1* have been reported to be associated with poor prognosis^{28–31}. Lipid metabolism and its metabolites can also induce tumor aggressiveness in LUAD³². Therefore, we extracted corresponding eight upregulated glycolytic genes and three downregulated lipid metabolic genes to characterize the metabolic reprogramming patterns of the samples and found LUAD to be clustered into three phenotypes (MPs). MP-III, characterized to have significantly high glycolytic and low lipid metabolic levels, was enriched in distant metastatic samples and associated with the poorest prognosis. Notably, in the TCGA dataset, there was no significant difference in stage I patients' OS between the high- and low-glycolysis groups clustered by eight glycolytic genes (log-rank $P = 0.66$, Supplementary Fig. 5a, b). While, in the high-glycolysis group, the patients with low expression patterns of lipid metabolism exhibited significantly shorter OS than the others (log-rank $P < 0.0001$, Supplementary Fig. 5c, d). These results suggested that high glycolytic and low lipid metabolic levels might synergistically accelerate tumor progression in LUAD. Additionally, we also found that a lot of other glycolytic and lipid metabolic genes were significantly correlated with the MPs (Pearson correlation, FDR < 0.05, Supplementary Data 2), such as a key glycolytic enzyme (*PKM*) gradually increasing from MP-I to

MP-III. Suggesting the stronger dysregulations of the two kinds of pathways in MP-III.

MP-III possessed the highest hypoxia, proliferation, and stemness scores, indicating its high metastatic potency. Notably, MP-III was characterized with the highest TMB but the lowest immune score. Previous studies have found that metabolic dysregulation of tumor cells could suppress the infiltration of immune cells and antitumor immunity³³, resulting in the high TMB tumor translating into an immunologically “cold” tumor^{34,35}. Our further genetic analysis revealed that the mutations of *TP53* and *KEAP1*, and deletions of *SETD2* and *PBRM1* might be underlying the metabolic reprogramming. The exploration of the genetic responsibility for MP-III would guide us in designing the studies of a clinical drug combination that may succeed in transforming these patients into those responsive to immunotherapy³⁵.

The MPs could be confirmed in multiple independent cohorts, and the stage I LUAD patients with MP-III exhibited the worst prognosis and the highest risk of developing distant metastasis after surgery. In contrast, MP-II was inclined to be enriched with lymph node metastatic samples, while it was not observed to have a significantly poorer prognosis than MP-I, which might be explained by the treatment strategy of lymph node dissection after surgery for LUAD³⁶. Notably, the independent cohorts were detected with microarray, scRNA-seq, and protein platforms, suggesting the cross-platform robustness of the MPs.

Metastasis imposes metabolic modifications that allow cancer cells to migrate away from primary tissues by altering their microenvironment³⁷. Therefore, we first used the scRNA-seq dataset of LUAD to confirm that MP-III was enriched in mBrain tumor cells and to demonstrate a metabolic evolutionary trajectory of epithelial cells from normal to MP-II and MP-III through MP-I. Notably, the evolutionary trajectory exhibited a unique low–low metabolic phenotype (State 4) in the scRNA-seq dataset, which has been reported in melanoma³⁴, meriting further validation and exploration. Subsequently, the intercellular communication analysis revealed the *ANGPTL* signaling pathway to be a unique one, in which MP-III population influences endothelial cells and fibroblasts. Padua et al. found that tumor cell-derived *ANGPTL4* disrupted vascular endothelial cell–cell junctions, increased the permeability of lung capillaries, and facilitated the trans-endothelial passage of breast tumor cells³⁸. Several studies have also shown the upstream role of *ANGPTL4* with respect to *VEGF*, indicating its role in angiogenesis^{39,40}. Our results also showed that all three MP populations could interact with

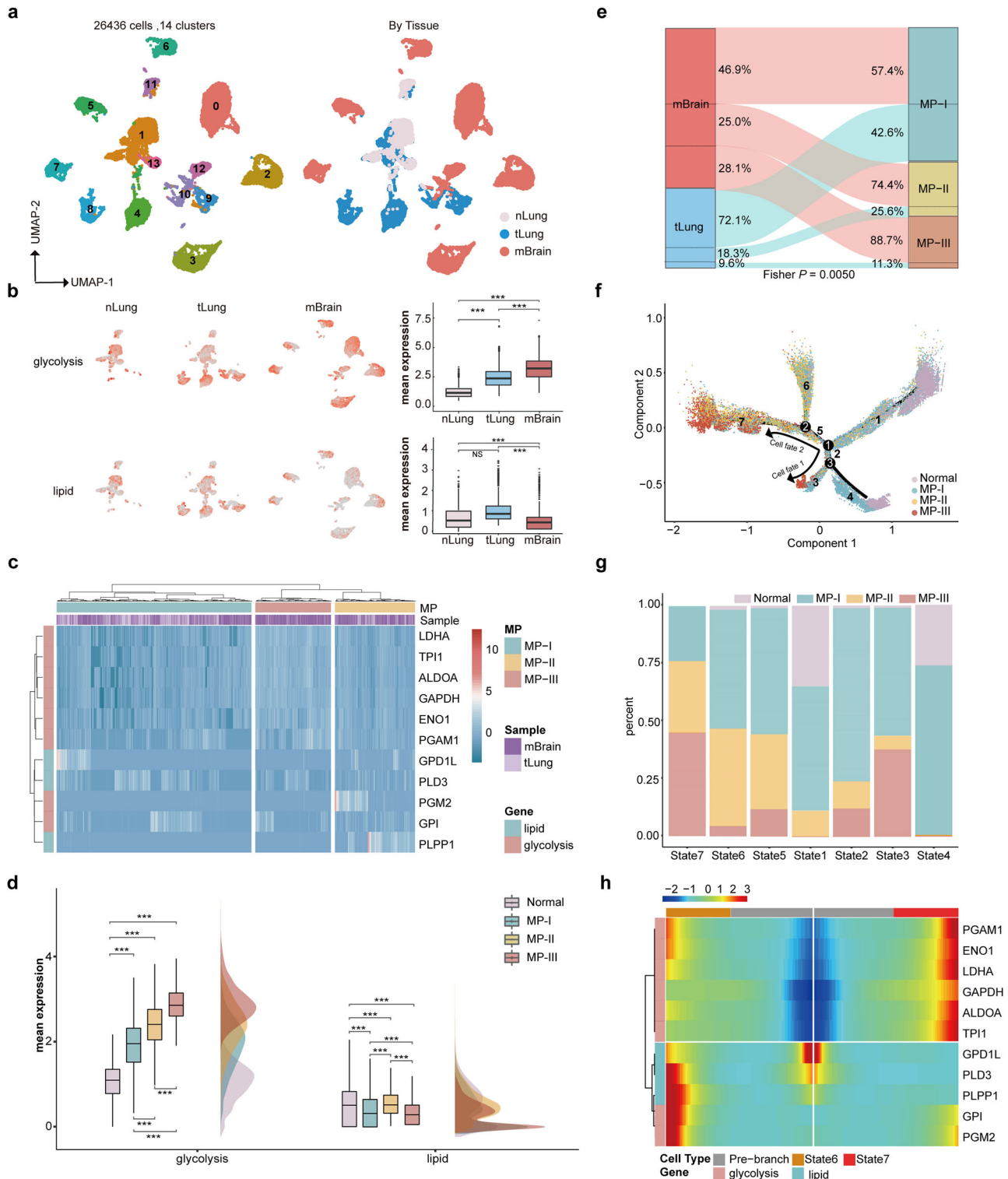


Fig. 5 Cell trajectory and branched expression analyses for the MPs in the single-cell RNA-sequencing dataset. **a** UMAP plot of 26,436 epithelial tumor cells colored based on the major cell lineages and normal, tLung, and mBrain tissues in the GSE131907 dataset. **b** Average expression of the eight glycolytic and three lipid metabolic genes within normal, tLung, and mBrain tissues represented by UMAP plots and box plots. **c** Hierarchical clustering heatmap for epithelial tumor cells based on the mRNA expression (Z-score) of the 11 metabolic genes. **d** Raincloud plot for the average expression of glycolytic and lipid metabolic genes in the three MPs and normal cells. **e** Sankey diagram showing the flow/change of tumor tissues (tLung/mBrain) in the MPs. **f** Unsupervised transcriptional trajectory of epithelial cells based on the 11 metabolic genes using Monocle (version 2), colored based on MP-I, -II, -III, and normal. **g** Relative percentages of the MP populations for each cell state as shown in trajectory. **h** The mRNA expression (Z-score) heatmap for the 11 metabolic genes in a branch-dependent manner. Genes (rows) are clustered and cells (columns) are ordered according to the pseudotime development, which along the pre-branch to State 6 and State 7. Boxplots extend from the 25th to 75th percentiles, the line indicates the median, and whiskers indicate the range.

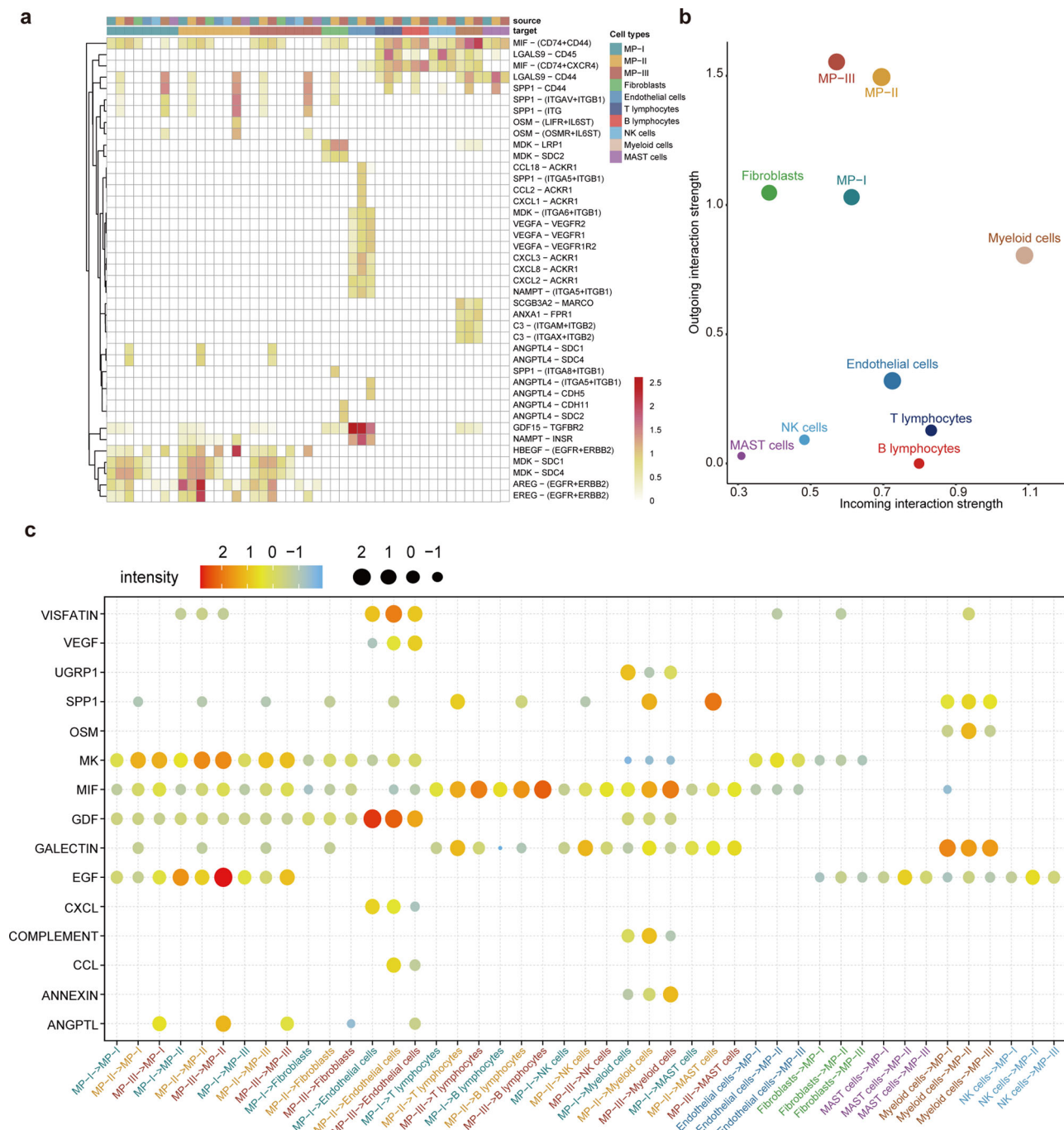


Fig. 6 Identification of ligand-receptor pairs across 10 cell populations. a Crosstalk intensity heatmap for significant ligand-receptor pairs across ten cell populations. **b** Dot plot showing the outgoing and incoming signaling patterns of intercellular communications for ten cell populations. **c** Crosstalk intensity heatmap for corresponding signaling pathways of ligand-receptor pairs in each cell population.

endothelial cells in the VEGF signaling pathway and that MP-III exhibited interactions with the most significant intensity. MP-III population not only contributed to its own EGF signaling pathway but also significantly activated the EGF pathway in the MP-II population, indicating the role of MP-III in MP-II deterioration. We also decoded more sophisticated interactome of the MPs with refined subdivided cell subtypes and observed similar results, such as ANGPTL and VEGF signaling pathways (Supplementary Fig. 6, Supplementary Data 3). We additionally observed that Cytotoxic and Exhausted CD8⁺ T lymphocytes be sources of *IFNG* in the IFN-II signaling pathway that acted on MP-II and MP-III populations (Supplementary Fig. 6). Finally, in order to

comprehensively elucidate the underlying molecular alterations for metastasis, we recapitulated a metabolic phenotype-specific cell-cell communication atlas of LUAD and linked it with cell-autonomous oncogenic events (*TP53*, *KEAP1*, *SETD2*, and *PBRM1*) through metabolic genes, thereby providing candidate therapeutic targets. Recent studies have shown that mutations in *TP53*^{41–43} and *KEAP1*⁴⁴ contribute to various metabolic disorders and play pivotal roles in metabolic reprogramming and cancer progression. Furthermore, the protein-protein interaction network from the STRING database (<https://string-db.org/>) indicated that *TP53* could activate the glycolysis pathway through up-regulating *PGM2*⁴¹, and hereby enhance the intercellular

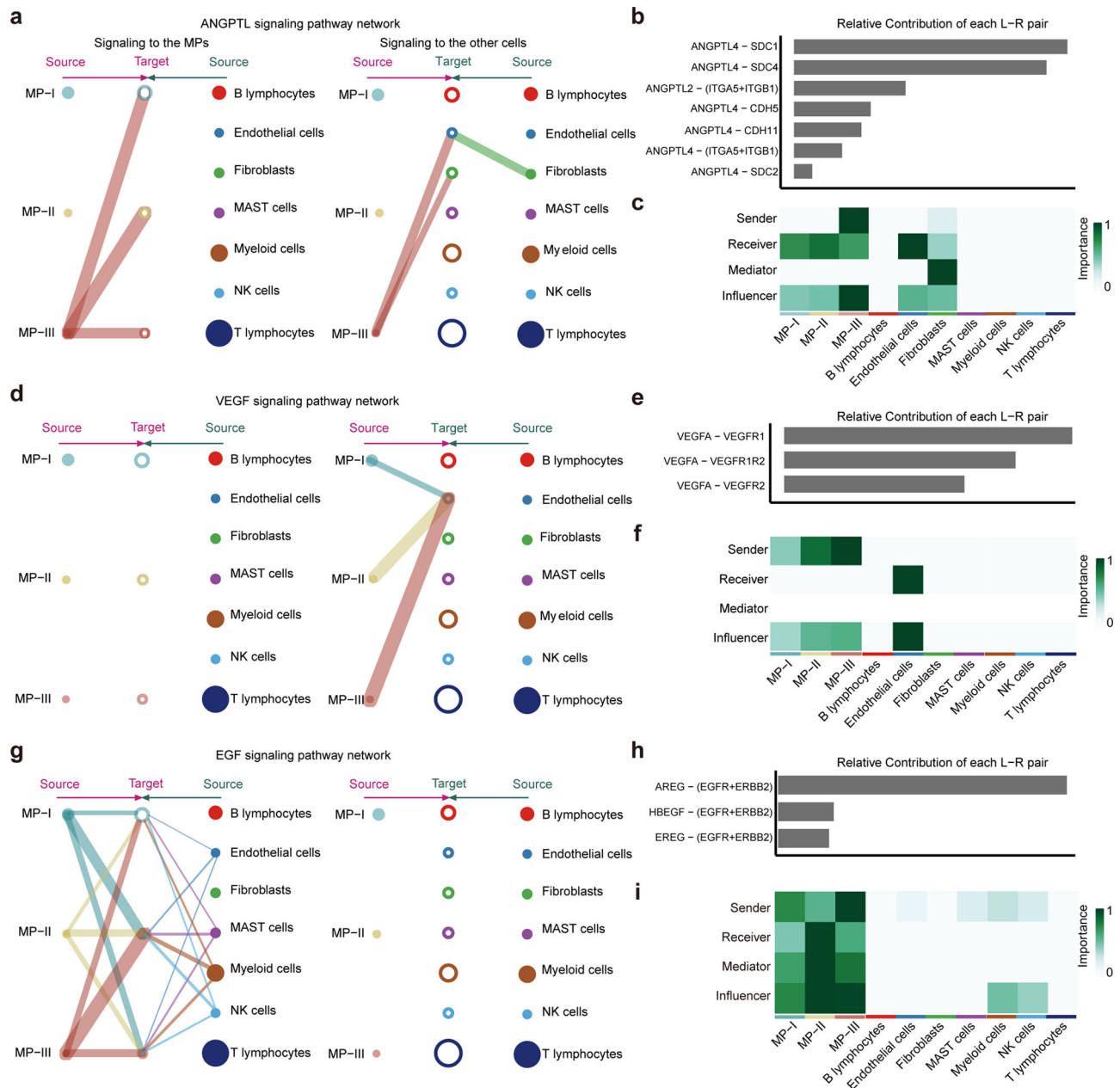


Fig. 7 Network centrality analysis of specific pathways. **a** Hierarchical plot showing the interactions between the MPs and other cells in the ANGPTL signaling pathway. This plot consists of two parts: Left and right portions highlight the autocrine and paracrine to MPs and to other cells, respectively. Solid and open circles represent source and target, respectively. Circle sizes are proportional to the number of cells in each cell group and edge width represents the communication probability. Edge colors are consistent with the signaling source. **b** Relative contribution of each ligand-receptor pair to the overall communication network of the ANGPTL signaling pathway. **c** Relative importance of each cell population based on the computed four network centrality measures of the ANGPTL signaling pathway. Details are described in the original article⁵⁷. **d-f** VEGF signaling pathway. **g-i** EGF signaling pathway.

communication of MP-III with endothelial cells on the VEGF signaling pathway (*VEGFA-KDR*). Meanwhile, *TP53* could regulate lipid metabolism by interacting with *HIPK4* to *PLD3*. Previous studies have shown that the mutation in *KEAP1* could activate *NRF2*, directing epithelial cells toward metabolic reprogramming of glucose metabolism^{44,45}. In this study, we found that *KEAP1* mutation might indirectly interact with a glycolytic gene (*LDHA*) through *COPS5*, which is related to the activation of glycolysis⁴⁶. Meanwhile, *KEAP1* might also affect lipid metabolism through indirect interaction with *TP53*. Additionally, the histone H3K36 methyltransferase *SETD2* and the chromatin remodeling factor *PBRM1* are frequently mutated or deleted in a variety of human tumors⁴⁷⁻⁵⁰. These studies indicated that

SETD2 and *PBRM1* deletions might participate in glycolysis-lipid metabolism discoordination by indirectly interacting with *TP53* and *KEAP1*.

This study still had some limitations. First, the metabolic phenotypes defined in this study were based on the mRNA expression of the 11 metabolic genes, thus the activities of glycolysis and lipid metabolisms need to be validated by estimating their metabolites. Second, the cell-cell interactions discovered in this study should be carefully determined, as they were speculative by the mRNA expression of ligands and their receptors. Though the significantly more interactions of the MP-III population with other populations were validated in another scRNA-seq dataset, the relative reliability of the communications still

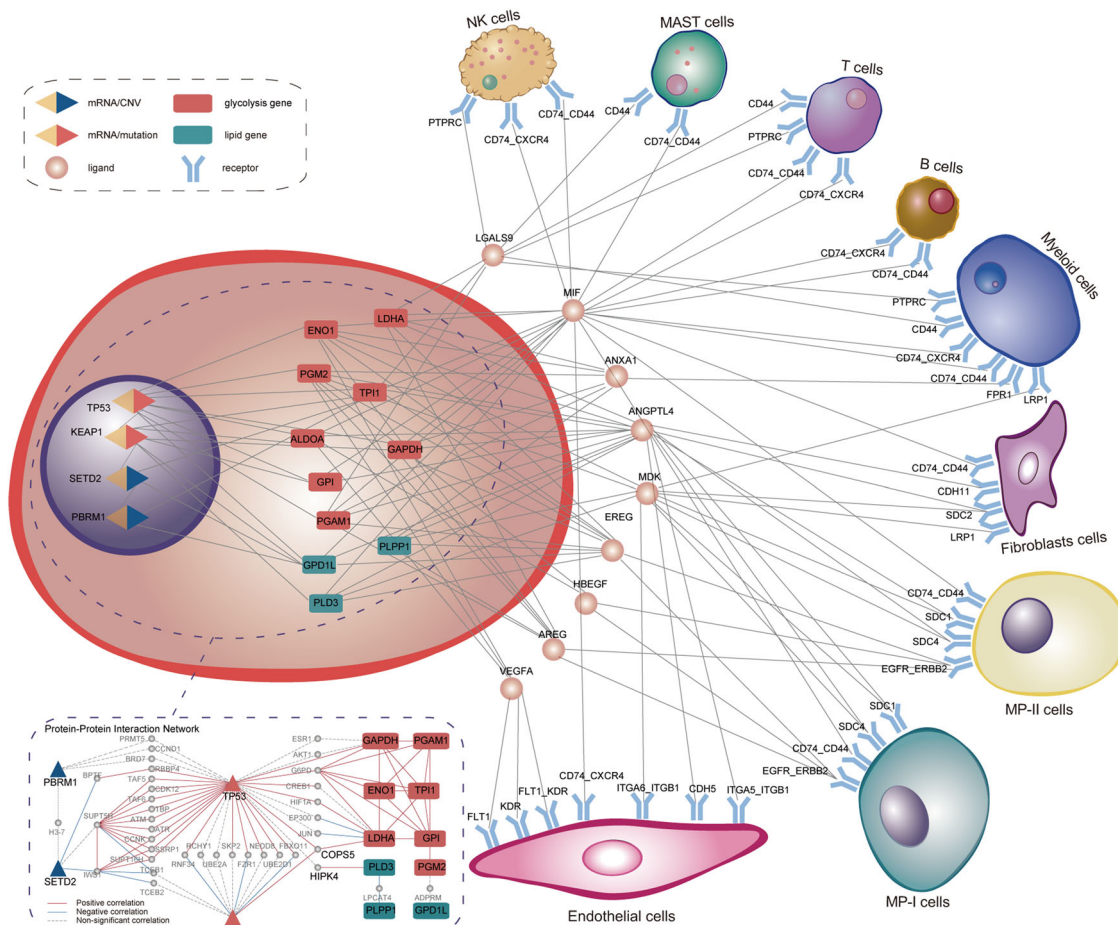


Fig. 8 Comprehensive network linking the intracellular oncogenic and metabolic events of MP-III with intercellular communications. The correlation network includes four genetic events (*TP53* mutation, *KEAP1* mutation, *SETD2* deletion, and *PBRM1* deletion), the 11 metabolic genes, and 45 ligand-receptor pairs of MP-III interacting with other cells in the tumor microenvironment. Here, the interacting genes expressed on MP-III were restricted to the nine ligands that were significantly upregulated in MP-III compared to MP-I. The nodes represent these genes and the edges represent the significant correlation between two nodes in the TCGA dataset (Pearson correlation, FDR < 0.05). The dotted box is the protein-protein interaction network between genetic events and the 11 metabolic genes from the STRING database, within the solid lines, represent significant correlation (Pearson correlation, $P < 0.05$).

needs further experimental verification. Furthermore, the other modes of metabolic reprogramming in altering the microenvironment, such as metabolites, merit further exploration.

In conclusion, we determined a metabolic phenotype (MP-III) relevant to the (occult) metastasis of LUAD, which was characterized by glycolytic and lipid metabolic discoordination. The Discovery of the four oncogenic events underlying this metabolic reprogramming and its pivotal interactions (especially the ANGPTL signaling pathway) with other cells provides candidate therapeutic strategies against LUAD metastasis.

Methods

Data sources. In this study, ten LUAD cohorts were obtained from the TCGA (<https://portal.gdc.cancer.gov/>), Gene Expression Omnibus (GEO, <https://www.ncbi.nlm.nih.gov/geo/>), and Clinical Proteomic Tumor Analysis Consortium data portal (CPTAC, <https://cptac-data-portal.georgetown.edu/>). Informed consent was obtained from all participants.

Two metastasis datasets (TCGA and GSE11969) that met the inclusion criteria of including LUAD patients with distant metastases, lymph node metastasis, and non-metastasis, were selected. The TCGA dataset included the mRNA expression profiles of 394 LUAD samples, comprising 25 distant metastatic, 160 lymph node metastatic samples (collectively termed “metastatic samples”), and 209 non-metastatic samples. Among them, mutation and CNV data were available for 390 and 385 samples, respectively. The GSE11969 dataset included the mRNA expression profiles of 18 distant metastatic, 15 lymph node metastatic, and 35 non-metastatic samples. Five other LUAD expression datasets for survival analysis were selected based on the

inclusion criteria that the number of stages I patients not receiving adjuvant therapy after surgical resection (treatment-naïve) was more than 50. Therein, the GSE13213 dataset also recorded the follow-up metastasis information of the stage I patients after surgery within five years, among which nine patients developed distant metastases, one patient had lymph node metastasis and 52 patients did not have any metastasis. Clinical information of the patients in the datasets is presented in Supplementary Table 5. Furthermore, we downloaded a paired mRNA and protein expression dataset containing 110 LUAD samples from CPTAC.

In addition, we collected two scRNA-seq LUAD datasets (GSE131907 and GSE123902). In the GSE131907 dataset, we extracted 208,506 cells derived from the primary tissues of 11 patients with tLung, metastatic tissues of 10 patients with mBrain, and 11 distant normal lung tissues. In the GSE123902 dataset, we extracted 18,511 cells derived from the primary tissues of seven patients with tLung.

Data pre-processing. For the RNA-seq dataset (TCGA) generated by the Illumina HiSeq 2000 platform, the fragments per kilobase of transcript per million mapped reads values were log₂-scaled plus 1 for gene expression measurements. For the scRNA-seq (GSE131907 and GSE123902) dataset derived from the Illumina HiSeq 2500 platform, the unique molecular identifier count for genes in each cell was log-normalized to transcripts per million-like values, and then log₂-scaled plus 1. Ensembl (ENSEM) gene IDs and symbols were mapped to their Entrez gene IDs. For the microarray datasets (GSE31210, GSE50081, and GSE68465) generated by Affymetrix platforms, a robust multi-array average algorithm⁵¹ was used to pre-process the raw data. For the microarray datasets (GSE42127 and GSE13213) generated by the Illumina and Agilent platforms, the originally processed data (series matrix files) were used. Probe IDs were mapped to gene IDs according to the corresponding platform files.

For somatic mutation data derived from the Illumina Genome Analyzer DNA Sequencing GAIx platform, only 17,821 nonsynonymous mutations were included, and a discrete mutation profile (mutated or not) was generated using Mutect2 somatic variant calls in mutation annotation format. CNV data were processed using the GISTIC algorithm⁵², with thresholds of 0.3 and -0.3 for amplified and deleted regions, respectively. CPTAC proteomic data was processed using the Common Data Analysis Pipeline⁵³.

Differential expression and functional enrichment analyses. The student's *t*-test was used to identify the DE genes between two groups. Functional enrichment analysis for DE genes was performed by a hypergeometric distribution model. Functional pathways were downloaded from the Kyoto Encyclopedia of Genes and Genomes signaling pathway database (<https://www.genome.jp/kegg/pathway.html>).

Identification of metabolic reprogramming phenotypes. First, hierarchical clustering was performed on samples to distinguish different clusters following Ward's method, which is based on Pearson correlation between the upregulated glycolytic and downregulated lipid metabolic genes. We pre-deleted the samples (or cells) whose expressions of all DE metabolic genes were zero before clustering and scaled the gene expression. Next, we calculated an MP-score estimating the expression imbalance of the glycolytic and lipid metabolic genes in each cluster as follows:

$$\text{MP-score} = \frac{\sum_{a=1}^n \sum_{i=1}^k \frac{G_{ai}}{n * k}}{\sum_{b=1}^m \sum_{i=1}^k \frac{L_{bi}}{m * k}} \quad (1)$$

where G_{ai} and L_{bi} represent the expression of *a*th glycolytic gene and *b*th lipid metabolic gene in *i*th sample, respectively; *n* and *m* are the numbers of glycolytic and lipid metabolic genes, respectively, and *k* represents the number of samples in the cluster.

Correlation and survival analyses. Wilcoxon rank test was used to determine differences in molecular scores (including hypoxia¹⁵, stemness¹⁶, proliferation¹⁷, and immune scores¹⁸) and TMB between multiple groups. Here, TMB was estimated as (total number of mutations in genes/total number of bases in genes × 10⁶). Fisher's exact test was performed to evaluate whether the percentages of stages, transcriptomics subtypes⁵⁴, and frequencies of genetic lesions were significantly different across the multiple groups. Pearson correlation analysis was used to estimate the correlation of gene expression with another gene expression, gene mutation statuses, CNVs, and protein expression, respectively.

OS was defined as the time from the date of initial surgical resection to the date of death or last contact (censored), which was truncated at 60 months. Survival curves were drawn using the Kaplan–Meier method and were statistically compared using log-rank test⁵⁵. Univariate Cox regression model was used to analyze the association between clinical factors of patients and their OS. HRs and 95% CIs were generated using Cox regression models.

Single-cell RNA-sequencing data analyses. The unsupervised clustering of epithelial cells was performed by “RunPCA” and “JackStraw” functions of Seurat package v4.0.1, based on the first 20 principal components of the top 2000 most variable genes among the whole-genome. UMAP (“RunUMAP” function) was used for the visualization of clustering. “Featureplot” function was used to show the average mRNA expression of genes on each cell. The subdivided cell subtypes were identified by SingleR v0.2.0 and manual validated in CellMarker database.

Cell trajectory and branched expression analyses were performed based on the Monocle package (version 2), using the reverse graph embedding machine learning algorithm to learn the changes in 11 metabolic gene expression sequences that each cell must go through as part of a dynamic biological process (here, metabolic reprogramming). The dimensionality of the cells was reduced by the DDRTree method, sequenced in pseudotime trajectory, and finally, visualized⁵⁶.

Inference of intercellular communications was analyzed using CellChat package v1.0.0⁵⁷. Ligand–receptor interactions and related signaling pathways were downloaded from the CellChat database (<https://www.cellchat.org/>). This method inferred the potential interaction intensity of a ligand–receptor pair between two cell populations, which considered their gene expression, signaling cofactors, and cell percentages⁵⁷. Significance was evaluated through a permutation test (1000 times). Only receptors and ligands expressed in more than 25% of cells in the specific cell subsets and those with significantly higher levels in specific cell populations than in other cells, with a threshold of $P < 0.05$ and log₂-fold change > 0.5, were analyzed. Outgoing and incoming strengths of each cell population were calculated as the cumulative interaction intensities of ligands and receptors expressed on this kind of cell respectively.

Statistics and reproducibility. The details about statistics used in different data analyses performed in this study are given in the respective sections of results and methods. Statistical analyses were performed using R version 3.4.0 (<https://www.r-project.org/>). *P*-values were adjusted using the Benjamini–Hochberg procedure for multiple testing to control FDR⁵⁸. Statistical significance was defined as two-sided $P < 0.05$ or FDR < 0.05 for multiple testing. Although there was no duplicate

sample, the experimental results were verified in multiple independent datasets, supporting the repeatability of the results.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

All datasets generated or analyzed during this study were collected from public databases, shown in Table 1 and Supplementary Table 5, which could be downloaded from the official website of the databases (see Methods). Data underlying figures in the main text are presented in Supplementary Data 4.

Code availability

Code utilized in these analyses is immediately available from the corresponding author upon reasonable request.

Received: 18 August 2021; Accepted: 7 February 2022;

Published online: 17 March 2022

References

- Siegel, R. L., Miller, K. D., Fuchs, H. E. & Jemal, A. Cancer Statistics, 2021. *CA Cancer J. Clin.* **71**, 7–33 (2021).
- Ettinger, D. S. et al. NCCN guidelines insights: non-small cell lung cancer, version 1.2020. *J. Natl Compr. Cancer Netw.* **17**, 1464–1472 (2019).
- Hawes, D., Neville, A. M. & Cote, R. J. Occult metastasis. *Biomed. Pharmacother.* **55**, 229–242 (2001).
- Yoshida, G. J. Metabolic reprogramming: the emerging concept and associated therapeutic strategies. *J. Exp. Clin. Cancer Res.* **34**, 111 (2015).
- Dang, C. V. Links between metabolism and cancer. *Genes Dev.* **26**, 877–890 (2012).
- Kim, J. & DeBerardinis, R. J. Mechanisms and implications of metabolic heterogeneity in cancer. *Cell Metab.* **30**, 434–446 (2019).
- Rieth, M. & Gold, D. D. Jr. Seasonal affective disorder: update 1988. *S. D. J. Med.* **41**, 17–20 (1988).
- Yang, J. et al. The enhancement of glycolysis regulates pancreatic cancer metastasis. *Cell Mol. Life Sci.* **77**, 305–321 (2020).
- Tanyi, J. L. et al. Role of decreased levels of lipid phosphate phosphatase-1 in accumulation of lysophosphatidic acid in ovarian cancer. *Clin. Cancer Res.* **9**, 3534–3545 (2003).
- Duan, L. et al. The role of GDF15 in bone metastasis of lung adenocarcinoma cells. *Oncol. Rep.* **41**, 2379–2388 (2019).
- Peng, X. et al. Molecular characterization and clinical relevance of metabolic expression subtypes in human cancers. *Cell Rep.* **23**, 255–269 e254 (2018).
- Karasinska, J. M. et al. Altered gene expression along the glycolysis-cholesterol synthesis axis is associated with outcome in pancreatic cancer. *Clin. Cancer Res.* **26**, 135–146 (2020).
- Takeuchi, T. et al. Expression profile-defined classification of lung adenocarcinoma shows close relationship with underlying major genetic changes and clinicopathologic behaviors. *J. Clin. Oncol.* **24**, 1679–1688 (2006).
- Hayes, D. N. et al. Gene expression profiling reveals reproducible human lung adenocarcinoma subtypes in multiple independent patient cohorts. *J. Clin. Oncol.* **24**, 5079–5090 (2006).
- Eustace, A. et al. A 26-gene hypoxia signature predicts benefit from hypoxia-modifying therapy in laryngeal cancer but not bladder cancer. *Clin. Cancer Res.* **19**, 4879–4888 (2013).
- Miranda, A. et al. Cancer stemness, intratumoral heterogeneity, and immune response across cancers. *Proc. Natl Acad. Sci. USA* **116**, 9020–9029 (2019).
- Whitfield, M. L., George, L. K., Grant, G. D. & Perou, C. M. Common markers of proliferation. *Nat. Rev. Cancer* **6**, 99–106 (2006).
- Yoshihara, K. et al. Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat. Commun.* **4**, 2612 (2013).
- Okayama, H. et al. Identification of genes upregulated in ALK-positive and EGFR/KRAS/ALK-negative lung adenocarcinomas. *Cancer Res.* **72**, 100–111 (2012).
- Der, S. D. et al. Validation of a histology-independent prognostic gene signature for early-stage, non-small-cell lung cancer including stage IA patients. *J. Thorac. Oncol.* **9**, 59–64 (2014).
- Tomida, S. et al. Relapse-related molecular signature in lung adenocarcinomas identifies patients with dismal prognosis. *J. Clin. Oncol.* **27**, 2793–2799 (2009).
- Tang, H. et al. A 12-gene set predicts survival benefits from adjuvant chemotherapy in non-small cell lung cancer patients. *Clin. Cancer Res.* **19**, 1577–1586 (2013).

23. Director's Challenge Consortium for the Molecular Classification of Lung, A. et al. Gene expression-based survival prediction in lung adenocarcinoma: a multi-site, blinded validation study. *Nat. Med.* **14**, 822–827 (2008).
24. Kim, N. et al. Single-cell RNA sequencing demonstrates the molecular and cellular reprogramming of metastatic lung adenocarcinoma. *Nat. Commun.* **11**, 2285 (2020).
25. Laughney, A. M. et al. Regenerative lineages and immune-mediated pruning in lung cancer metastasis. *Nat. Med.* **26**, 259–269 (2020).
26. Hanahan, D. & Weinberg, R. A. Hallmarks of cancer: the next generation. *Cell* **144**, 646–674 (2011).
27. Ye, T., Liang, Y., Zhang, D. & Zhang, X. MicroRNA-16-1-3p represses breast tumor growth and metastasis by inhibiting PGK1-mediated Warburg effect. *Front. Cell Dev. Biol.* **8**, 615154 (2020).
28. Chang, Y. C. et al. Feedback regulation of ALDOA activates the HIF-1 α /MMP9 axis to promote lung cancer progression. *Cancer Lett.* **403**, 28–36 (2017).
29. Puzone, R. et al. Glyceraldehyde-3-phosphate dehydrogenase gene over expression correlates with poor prognosis in non small cell lung cancer patients. *Mol. Cancer* **12**, 97 (2013).
30. Hou, X. M., Yuan, S. Q., Zhao, D., Liu, X. J. & Wu, X. A. LDH-A promotes malignant behavior via activation of epithelial-to-mesenchymal transition in lung adenocarcinoma. *Biosci. Rep.* <https://doi.org/10.1042/BSR20181476> (2019).
31. Sun, Q. et al. Phosphoglyceric acid mutase-1 contributes to oncogenic mTOR-mediated tumor growth and confers non-small cell lung cancer patients with poor prognosis. *Cell Death Differ.* **25**, 1160–1173 (2018).
32. Nema, R., Shrivastava, A. & Kumar, A. Prognostic role of lipid phosphate phosphatases in non-smoker, lung adenocarcinoma patients. *Comput. Biol. Med.* **129**, 104141 (2021).
33. Ramapriyan, R. et al. Altered cancer metabolism in mechanisms of immunotherapy resistance. *Pharm. Ther.* **195**, 162–171 (2019).
34. Spranger, S., Bao, R. & Gajewski, T. F. Melanoma-intrinsic beta-catenin signalling prevents anti-tumour immunity. *Nature* **523**, 231–235 (2015).
35. Parra, E. R. et al. Effect of neoadjuvant chemotherapy on the immune microenvironment in non-small cell lung carcinomas as determined by multiplex immunofluorescence and image analysis approaches. *J. Immunother. Cancer* **6**, 48 (2018).
36. Lu, Y., Ma, T., Wang, L. & Xue, T. Advances in lymph node metastasis and lymph node dissection in early non-small cell lung cancer. *Zhongguo Fei Ai Za Zhi* **22**, 520–525 (2019).
37. Alvarez-Elizondo, M. B. et al. Micropatterned topographies reveal measurable differences between cancer and benign cells. *Med. Eng. Phys.* **75**, 5–12 (2020).
38. Padua, D. et al. TGF β primes breast tumors for lung metastasis seeding through angiopoietin-like 4. *Cell* **133**, 66–77 (2008).
39. Lu, Q., Lu, P., Chen, W., Lu, L. & Zheng, Z. ANGPTL-4 induces diabetic retinal inflammation by activating Proflin-1. *Exp. Eye Res.* **166**, 140–150 (2018).
40. Yang, X., Cheng, Y. & Su, G. A review of the multifunctionality of angiopoietin-like 4 in eye disease. *Biosci. Rep.* <https://doi.org/10.1042/BSR20180557> (2018).
41. Kondoh, H. et al. Glycolytic enzymes can modulate cellular life span. *Cancer Res.* **65**, 177–185 (2005).
42. Liu, J., Zhang, C., Hu, W. & Feng, Z. Tumor suppressor p53 and its mutants in cancer metabolism. *Cancer Lett.* **356**, 197–203 (2015).
43. Lacroix, M., Riscal, R., Arena, G., Linares, L. K. & Le Cam, L. Metabolic functions of the tumor suppressor p53: Implications in normal physiology, metabolic disorders, and cancer. *Mol. Metab.* **33**, 2–22 (2020).
44. Singh, A. et al. Transcription factor NRF2 regulates miR-1 and miR-206 to drive tumorigenesis. *J. Clin. Invest.* **123**, 2921–2934 (2013).
45. Haley, J. A. et al. Decoupling of Nrf2 expression promotes mesenchymal state maintenance in non-small cell lung cancer. *Cancers* <https://doi.org/10.3390/cancers11101488> (2019).
46. Li, P. et al. Cops5 safeguards genomic stability of embryonic stem cells through regulating cellular metabolism and DNA repair. *Proc. Natl Acad. Sci. USA* **117**, 2519–2525 (2020).
47. Yuan, H. et al. Histone methyltransferase SETD2 modulates alternative splicing to inhibit intestinal tumorigenesis. *J. Clin. Invest.* **127**, 3375–3391 (2017).
48. Zhou, Y. et al. Histone methyltransferase SETD2 inhibits tumor growth via suppressing CXCL1-mediated activation of cell cycle in lung adenocarcinoma. *Aging* **12**, 25189–25206 (2020).
49. Trpkov, K. et al. New developments in existing WHO entities and evolving molecular concepts: The Genitourinary Pathology Society (GUPS) update on renal neoplasia. *Mod. Pathol.* **34**, 1392–1424 (2021).
50. Shmakova, A., Frost, M., Batie, M., Kenneth, N. S. & Rocha, S. PBRM1 cooperates with YTHDF2 to control HIF-1 α protein translation. *Cells* <https://doi.org/10.3390/cells10061425> (2021).
51. Irizarry, R. A. et al. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* **4**, 249–264 (2003).
52. Mermel, C. H. et al. GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. *Genome Biol.* **12**, R41 (2011).
53. Rudnick, P. A. et al. A description of the clinical proteomic tumor analysis consortium (CPTAC) common data analysis pipeline. *J. Proteome Res.* **15**, 1023–1032 (2016).
54. Wu, D. et al. Gene-expression data integration to squamous cell lung cancer subtypes reveals drug sensitivity. *Br. J. Cancer* **109**, 1599–1608 (2013).
55. Bland, J. M. & Altman, D. G. The logrank test. *Br. Med. J.* **328**, 1073 (2004).
56. Qiu, X. et al. Reversed graph embedding resolves complex single-cell trajectories. *Nat. Methods* **14**, 979–982 (2017).
57. Jin, S. et al. Inference and analysis of cell-cell communication using CellChat. *Nat. Commun.* **12**, 1088 (2021).
58. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B* **57**, 289–300 (1995).

Acknowledgements

This study was supported by grants from the National Natural Science Foundation of China (81872396 to L.Q., 61873075, 32070673 to X.L.); the National Key R&D Program of China (2018YFC2000100); the Postdoctoral Scientific Research Developmental Fund (grant number LBH-Q16166 to Y.G.); Heilongjiang Touyan Innovation Team Program; National Undergraduate Innovation and Entrepreneurship Training Program (202110226018 to L.Q.).

Author contributions

X.L. (Xia Li) conceived the idea, L.S.Q. designed the experiments and modified the paper, X.L. (Xin Li), L.F.T. designed and performed the experiments, wrote the paper, J.X.D., X.Y.Q. and J.X.Z. analyzed the gene expression data, H.T.Q., M.Y.L. and Y.X.L. pre-processed the radiogenomic data, W.Y.Z. and Y.Y.G. understanding the results. All authors approved the final version.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s42003-022-03135-z>.

Correspondence and requests for materials should be addressed to Lishuang Qi or Xia Li.

Peer review information *Communications Biology* thanks the anonymous reviewers for their contribution to the peer review of this work. Primary Handling Editor: George Inglis. Peer reviewer reports are available.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022