Clinical Significance and Regulatory Network of Mast Cells in Early-Stage Lung Adenocarcinoma

Zhen Wang

Sun Yat-Sen University Cancer Prevention and Treatment Center: Sun Yat-sen University Cancer Center

Bin Luo

Sun Yat-Sen University Cancer Prevention and Treatment Center: Sun Yat-sen University Cancer Center

Pengfei Duan

Sun Yat-Sen University Cancer Prevention and Treatment Center: Sun Yat-sen University Cancer Center

Han Yang

Sun Yat-Sen University Cancer Prevention and Treatment Center: Sun Yat-sen University Cancer Center

Peng Lin (✉ linpeng@sysucc.org.cn)

Sun Yat-Sen University Cancer Prevention and Treatment Center: Sun Yat-sen University Cancer Center

Research Article

Keywords: mast cell, early-stage lung adenocarcinoma, upstream regulator, immune infiltration, prognosis.

Posted Date: November 15th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-1055481/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Title: Clinical significance and regulatory network of mast cells in early-stage lung adenocarcinoma

Authors: Zhen Wang^{1#}, Bin Luo^{1#}, Pengfei Duan¹, Han Yang^{1*}, Peng Lin^{1*}.

[#] Zhen Wang and Bin Luo make an equal contribution to this work.

* Corresponding author.

Affiliations:

¹ Department of Thoracic Oncology, State Key Laboratory of Oncology in South China, Collaborative Innovation Center for Cancer Medicine, Sun Yat-sen University Cancer Center, Guangzhou 510060, People's Republic of China.

Correspondence to: Han Yang and Peng Lin

Address: Department of Thoracic Oncology, State Key Laboratory of Oncology in South China, Collaborative Innovation Center for Cancer Medicine, Sun Yat-sen University Cancer Center, 651 Dongfengdong, Guangzhou, Guangdong 510060, China.

Email of Corresponding author: Han Yang, yanghan@sysucc.org.cn

Peng Lin, linpeng@sysucc.org.cn

1 **ABSTRACT**

2 **Background**

3 The role and upstream regulation of mast cells (a crucial constituent of the tumor
4 immune microenvironment) in early-stage lung adenocarcinoma (LUAD) remains
5 unclear.

6 **Methods**

7 Using CIBERSORT, we analyzed the expression profile downloaded from the NCBI
8 Gene Expression Omnibus (GEO) database. Next, using weighted gene coexpression
9 network analysis (WGCNA) and enrichment analysis, we identified upstream
10 regulators of mast cells. Finally, we verified their relevance using data from The Cancer
11 Genome Atlas (TCGA) and Genotype-Tissue Expression (GTEx) databases.

12 **Results**

13 Among patients with early-stage LUAD, those with high mast cell infiltration levels
14 had a better prognosis, and mast cell infiltration levels were closely associated with the
15 mutation status of TP53, STK11 and EGFR. CAMK2D, AURKA, AURKB, PLK1 and
16 EZH2 showed strong correlations with mast cell infiltration and a key growth factor of
17 mast cells (KITLG), and these 5 regulators were closely correlated with the survival of
18 patients with early-stage lung adenocarcinoma.

19 **Conclusion**

20 Mast cells play a protective role in patients with early-stage LUAD. We speculate that
21 CAMK2D, AURKA, AURKB, PLK1 and EZH2 promote or inhibit the infiltration of
22 mast cells in early-stage lung adenocarcinoma tissues by regulating KITLG, likely

1 serving as novel targets in immunotherapy for LUAD.

2 **Keywords:** mast cell; early-stage lung adenocarcinoma; upstream regulator; immune
3 infiltration; prognosis.

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

25

26

27

28

1 INTRODUCTION

2 Lung cancer remains the malignancy with the highest morbidity and mortality
3 worldwide, of which adenocarcinoma is the most common histotype (1). As the most
4 prevalent subtype of lung cancer, lung adenocarcinoma (LUAD) has been increasing in
5 smokers and even nonsmokers (2). Fortunately, with the popularity of lung cancer
6 screening, an increasing number of lung cancers, particularly lung adenocarcinomas,
7 are screened out at an early stage (3-5). Patients with early-stage LUAD usually receive
8 surgery, chemotherapy and radiotherapy; however, in recent years, immunotherapy has
9 received increasing attention in LUAD treatment (6, 7). However, considering the
10 complexity of the tumor immune microenvironment, more work is needed to elucidate
11 the LUAD immune microenvironment.

12 As an important component of the human immune system, mast cells are widely
13 distributed, particularly in parts of the boundary between the external environment and
14 body, such as the respiratory system and digestive and urinary tracts (8, 9). Human mast
15 cells are derived from specific bone marrow hematopoietic stem cells, CD34+ and
16 CD117+ (10). Then mast cells migrate into tissues and mature, relying on the
17 surrounding microenvironment (11, 12). In the respiratory system, progenitors of mast
18 cells migrate from blood into lung tissues and acquire tissue-specific phenotypes (13).
19 They can then participate in maintaining the stability of the lung microenvironment and
20 contribute to innate and adaptive immunity in the respiratory system (14).

21 During tumorigenesis and progression, mast cells play a complicated role. They
22 either secrete pro-tumorigenic molecules, such as angiogenic and lymphangiogenic
23 factors, or secrete antitumorigenic molecules, such as TNF- α and IL-9 (15, 16).
24 Therefore, mast cells play different roles in different carcinoma types, even at different
25 stages of the same type of carcinoma. In this study, we focused on the effect of mast
26 cells in patients with early-stage LUAD (stage I and stage II) and constructed a
27 regulatory network of mast cell infiltration using a series of bioinformatics analysis
28 methods.

METHODS

Date source and processing

All the data used in our study were from public databases, including Gene Expression Omnibus (GEO) and The Cancer Genome Atlas (TCGA) (17, 18). The gene expression profile of GSE72094 was searched and downloaded from the Gene Expression Omnibus database (<http://www.ncbi.nlm.nih.gov/geo/>). GSE72094 comprises 442 lung adenocarcinoma samples, which were profiled on gene expression arrays. The prevalent driver mutations (KRAS, EGFR) and tumor suppressor mutations (STK11 and TP53) were also assessed (19).

Probes in this dataset were annotated using gene symbol information from soft formatted family files, the probe annotated using multiple genes was deleted, and the average value was recorded as the final expression value for the gene matched by multiple probes.

The transcriptome and clinical data of early-stage (stage I and stage II) LUAD in TCGA were downloaded from the National Cancer Institute's (NCI's) Genomic Data Commons (GDC) (<https://portal.gdc.cancer.gov/>) (20).

CIBERSORT estimation

The immunocyte infiltration levels of all samples were calculated using CIBERSORT in R studio with permutations set to 100. Using the expression signatures of 547 gene markers and a deconvolution algorithm, CIBERSORT can quantify the relative proportions of 22 types of immunocytes (21). Samples with a CIBERSORT result of $p < 0.05$ were reserved for subsequent analysis.

Immunocytes derived from CIBERSORT contain both resting mast cells and activated mast cells. In our study, the mast cell infiltration level was defined as the proportion of resting mast cells plus that of activated mast cells.

Survival and clinicopathologic characteristic analysis

Early-stage samples (stage I and stage II) were selected from the samples using a CIBERSORT output of $p < 0.05$. First, the Mann-Whitney test was performed to assess whether a difference existed in the degree of mast cell infiltration between stage I and

stage II patients. Next, the selected samples were divided into high infiltration and low infiltration groups in terms of the median mast cell proportions or upper quartile of mast cell proportions. Finally, Kaplan-Meier plots were generated to analyze whether a difference existed in overall survival (OS) between the groups. The relevant data were analyzed and plotted using the “ggpubr”, “survival” and “survminer” packages in R studio (22-24).

All the patients were divided into two groups according to the median mast cell proportions, and then the clinicopathological differences between the groups were analyzed by the Wilcoxon rank-sum test and Pearson's chi-squared test.

Weighted gene co-expression network analysis (WGCNA)

The transcriptome data of selected early-stage samples in GSE72094 were further estimated using WGCNA to detect genes related to mast cell infiltration. Weighted gene co-expression network analysis (WGCNA) is a bioinformatic algorithm that calculates highly correlated genes and assesses relevance between gene sets and clinicopathological traits (25).

The whole calculation process was completed using the “WGCNA” package in R. First, the top 5,000 genes were selected according to average expression. Second, the power parameter was calculated, and the appropriate soft-thresholding power for network construction was selected. Third, using a dynamic tree-cutting algorithm, module eigengenes (MEs) comprising a list of co-expressed genes were obtained. Finally, the correlation between MEs and clinicopathologic traits was assessed using Pearson's correlation. The ME displaying the most significant correlation with a related trait was selected for further analysis.

Upstream regulators of module eigengenes (MEs)

Expression2Kinases (X2K) contains a mass of transcription factors and kinase libraries that can computationally predict upstream regulatory networks from the signatures of co-expressed genes in related MEs (26). The installer for X2K was downloaded from: <http://www.maayanlab.net/X2K>. First, using promoter analysis of CHIP-X enrichment analysis, we identified the upstream transcription factors (TFs) of

the gene set contained in related MEs (27). Upstream protein kinases were then detected by kinase enrichment analysis (28). The network, which contains the TFs, protein kinases, and intermediate proteins of co-expressed genes, was mapped using Cytoscape software (29).

Correlation analysis

Stem cell factor (SCF), also known as KITLG, is an essential growth factor for the proliferation, differentiation and apoptosis suppression of mast cells (14, 30, 31). KITLG acts through binding to a receptor tyrosine kinase (RTK), which is known as KIT (CD117) (14, 31, 32). Therefore, we next calculated the correlations across TFs or kinases, KITLG, KIT and mast cell infiltration levels, which were processed and visualized using the R package “corrplot” (33).

In addition to the respiratory system, mast cells are also widely distributed in the urinary tract and digestive tract. Thus, we further analyzed the correlation between related upstream regulators and KITLG in these normal tissues. Gene Expression Profiling Interactive Analysis (GEPIA) is a website for cancer and normal gene expression profiling and interactive analyses, and Genotype-Tissue Expression (GTEx) is a project covering genetic variation and gene expression in human normal tissues (34, 35). Using the data from the GTEx database, we analyzed these correlations on the GEPIA website.

Survival analysis of KITLG and upstream regulators

All the samples of GSE72094 were divided into two groups in terms of the median expression of these genes. Kaplan-Meier plots were generated to analyze whether a difference existed in overall survival (OS) between the groups. The relevant data were also analyzed and plotted using the “ggpubr”, “survival” and “survminer” packages in R studio (22-24).

RESULTS

Patients with early-stage LUAD benefit from high mast cell infiltration

The proportions of 22 immunocytes in samples with a CIBERSORT result of $p < 0.05$ were exhibited in a cumulative histogram, where different colors represented diverse cell subsets (Fig. 1A). Considering that the early-stage LUAD in our study included both stage I and stage II samples, we first tested whether a difference existed in the mast cell proportions between the groups. No significant difference was found between the groups (Fig. 1B). Next, via Kaplan-Meier survival analysis, we observed that the survival of patients in the high mast cell infiltration group was significantly better than that in the low infiltration group when using the upper quartile as the cutoff value (Fig. 1c; $p = 0.012$).

To validate this phenomenon, we used data from TCGA for survival analysis. Although the P value did not reach a significant difference, patients in the high mast cell infiltration group still survived better than those in the low infiltration group (Fig. 4A). The correlation between T stage, N stage and survival status of patients stratified by mast cell infiltration levels is presented in a Sankey plot (Fig. 4B).

Correlation between mast cell infiltration and clinicopathological characteristics

The detailed clinicopathological characteristics of 309 patients with early-stage LUAD in GSE72094 are shown in Table 1. The 309 patients were divided into two groups according to the median value of mast cell proportions, with 154 patients in the high-infiltration group and 155 patients in the low-infiltration group. Between the groups, significant differences were found in sex ratio ($p = 0.034$), EGFR mutation status ($p = 0.032$), and TP53 mutation status ($p < 0.001$). The difference in SKT11 mutation status between the groups was also approximately statistically significant ($p = 0.058$).

The Sankey plot graphically shows the relationship across the tumor stage, TP53 mutation status and EFGG mutation status of patients stratified by mast cell infiltration levels (Fig. 1D). As shown in our results, the patients with a higher mast cell infiltration level have a lower TP53 mutation rate, while a higher EGFR mutation rate.

1 Identification of gene sets associated with mast cell infiltration

2 Optimal soft-thresholding power is essential for WGCNA. In the analysis of
3 GSE72094, when the soft-thresholding power is 5, a scale-free co-expression network
4 can be constructed. Next, using a dynamic pruning method, the top 5,000 genes were
5 clustered into 12 modules. Interactive relationships between 12 modules and
6 clinicopathological traits are visualized as a network heatmap (Fig. 2A). Correlation
7 analysis showed that the brown module had the highest negative correlation coefficient
8 with the mast cell proportions (R : -0.43), and the red module had the highest positive
9 correlation coefficient (R : 0.34). The gene significance vs. module membership for
10 these two modules are shown as scatter plots (Fig. 2B, C).

11 Upstream regulators of mast cells

12 The correlation coefficient between the mast cell proportions and brown module
13 was -0.43, indicating that genes in this module may be involved in inhibiting the
14 infiltration of mast cells in LUAD tissues. Similarly, the genes in the red module may
15 be involved in promoting mast cell infiltration. Using enrichment analysis of
16 Expression2Kinases, we found upstream regulators of these genes, including
17 transcription factors (TFs) and kinases. The top 10 transcription factors enriched by
18 genes in the brown module are E2F4, MYCN, SUZ12, SOX2, MTF2, KDM5B,
19 FOXA2, NANOG, E2F1, and YAP1. The top 10 kinases enriched by genes in the brown
20 module are CDK1, PLK1, AURKA, CHEK1, AURKB, CHEK2, CDK2, ATR, MAPK1,
21 and CDK3. The top 10 transcription factors enriched by genes in the red module are
22 SUZ12, MTF2, RNF2, EZH2, JARID2, EED, PHC1, SOX2, PAX3, and AR. The top
23 10 kinases enriched by genes in the red module are PRKG1, ACVR2A, PRKACA,
24 BMPR1A, IGF1R, TEK, CAMK2D, PRKACB, BMPR2, and TEC. The network
25 containing these transcription factors, intermediate proteins and kinases is shown in
26 figure 3A and figure 3D.

27 Correlations across mast cells, KITLG, KIT and upstream regulators

28 A correlation heat map can vividly reflect how the variates are associated with each
29 other. From the correlation heat map across TFs or kinases, KITLG, KIT and mast cell

infiltration levels, we identified upstream regulators associated with mast cell infiltration. TFs or kinases with a correlation coefficient greater than 0.4 or less than -0.4 were considered to have a strong correlation. By analyzing transcriptome data of GSE72094, AURKA, AURKB, PLK1 and EZH2 were identified to have strongly negative correlations with mast cell infiltration levels (Fig. 3C, E). These upstream regulators also show obviously negative correlations with KITLG or KIT (Fig. 3C, E).

Next, the data from TCGA were used to verify these correlations. Although the correlations between these 4 upstream regulators and mast cell infiltration levels did not reach -0.4, the correlations across them were still statistically significant and relatively obvious (Fig. 4D, E). Furthermore, in the analysis of transcriptome data from TCGA, CAMK2D showed a strongly positive correlation with the mast cell infiltration level (R : 0.41; Fig. 4F), and this correlation was also statistically significant in the analysis of GSE72094 (R : 0.28; Fig. 3F). Similarly, an obvious correlation existed between CAMK2D and KITLG (R : 0.37; Fig. 4F). The potential interactions of tumor cells, mast cells, KITLG and related upstream regulators in early-stage LUAD tissues is shown in figure 6.

Correlations between KITLG and upstream regulators in normal tissues

Correlations between KITLG and upstream regulators in the urinary tract and digestive tract are shown in figure 5 and table 2. CAMK2D and KITLG show significant positive correlations in these normal human tissues, particularly in bladder, transverse colon and stomach tissues. AURKA, AURKB, PLK1 and EZH2 also show significant negative correlations with KITLG in these tissues, particularly in the transverse colon and terminal ileum.

KITLG and upstream regulators predict the prognosis of patients with early-stage LUAD

As an indispensable growth factor of mast cells, early-stage lung adenocarcinoma patients with high expression of KITLG also have a better prognosis (Fig. 7A; p = 0.0025). CAMK2D, as an upstream regulator positively correlated with mast cell infiltration, showed high expression, which also indicated a better prognosis in patients

1 with early-stage lung adenocarcinoma (Fig. 7B; $p = 0.00014$). By contrast, as upstream
2 regulators negatively related to mast cell infiltration, AURKA, AURKB, PLK1 and
3 EZH2 showed a worse prognosis at higher expression levels (Fig. 7C, D, E, F).

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

25

26

27

28

29

DISCUSSION

Although many previous studies have focused on the immune microenvironment and immunotherapy of lung cancer, few reports have investigated the role of mast cells in lung cancer. Whether in primary or secondary lung carcinoma, the tumor microenvironment (TME) contains abundant therapeutic targets for immunotherapy (36). As an important component of the human immune system, the potential value of mast cells in immunotherapy cannot be ignored.

In our study, among patients with early-stage lung adenocarcinoma, those with a high proportion of mast cells had obvious survival advantages. Because mast cells play a double-sided role in the development of tumors, previous reports on the prognostic significance of mast cells on tumors are often contradictory (37, 38). Using immunohistochemical staining with monoclonal antibodies for specific markers in mast cells, Takanami et al found that lung adenocarcinoma patients with high mast cell infiltration levels show a poor prognosis and believed that the promotion of mast cells on angiogenesis caused this phenomenon (39). Considering that the stages of patients recruited in this study contained stage I to stage III, the author did not explain the role of mast cells in patients with early-stage LUAD. Similarly, using immunohistochemical staining with an antibody for Kit (CD-117) in 4,444 samples, Rajput et al detected that the mast cell infiltration level is a protective factor in patients with invasive breast cancers (40). Additionally, mast cells were shown to inhibit tumor progression by secreting various tumor cytotoxic cytokines and promoting lymphocyte infiltration (41-43). A study similar to ours reached the same conclusion. Using xCell, a different computational cell-type quantification method, Bao et al estimated the abundance of different immunocytes of multiple transcriptome files and found that high mast cell abundance benefits the survival of patients with early-stage LUAD (44, 45). Therefore, we believe that, in patients with early-stage LUAD, mast cells play an antitumor role by secreting cytotoxic cytokines and recruiting antitumor immunocytes.

TP53 and STK11 (serine/threonine kinase 11) are two common tumor suppressor genes, whose mutations can lead to the occurrence and progression of various tumors,

1 including lung adenocarcinoma (46, 47). In our study, the proportions of TP53 and
2 STK11 mutations in patients in the high mast cell infiltration group were lower than
3 those in patients in the low mast cell infiltration group, a finding that is consistent with
4 their protective effect on patients with lung adenocarcinoma. EGFR (epidermal growth
5 factor receptor) mutations occur in approximately 16% of patients with non-squamous
6 NSCLC (non-small cell lung cancer). Additionally, as the most widely known
7 therapeutic target in lung cancer, EGFR inhibitors have shown survival benefits in
8 millions of patients worldwide (48, 49). In our study, the patients with high mast cell
9 infiltration levels had higher EGFR mutation proportions than those of the other group,
10 suggesting that patients with higher mast cell infiltration levels are more likely to
11 benefit from EGFR-TKI (EGFR tyrosine kinase inhibitor) treatment. Biton et al
12 reported that the combined use of TP53, STK11 and EGFR effectively predicts the
13 expression of programmed cell death ligand-1 (PD-L1) by malignant tumor cells, and
14 patients with TP53-Mut/STK11-EGFR-WT tumors achieve better progression-free
15 survival (PFS) from anti-PD-1/PD-L1 treatment (50). Therefore, we believe that the
16 infiltration proportions of mast cells in lung adenocarcinoma tissue are associated with
17 mutations in multiple oncogenes/tumor suppressor genes and can be used as a predictor
18 of efficacy in patients receiving anti-PD-1/PD-L1 treatment.

19 CAMK2D (calcium/calmodulin-dependent protein kinase II delta) is a protein-
20 coding gene, and its product belongs to the serine/threonine protein kinase family and
21 Ca (2+)/calmodulin-dependent protein kinase subfamily (51). The pathways in which
22 CAMK2D participates include the aldosterone synthesis and secretion and immune
23 response IFN gamma signaling pathways (52). A previous study confirmed that IFN- γ
24 regulates mast cell survival and that this cytokine induces mast cell death in vitro and
25 in vivo (53). Combined with our study on the correlation between CAMK2D and
26 KITLG in lung cancer tissues and normal tissues of the digestive and urinary tracts, we
27 speculate that CAMK2D regulates KITLG and IFN- γ to promote the infiltration of mast
28 cells in early-stage lung adenocarcinoma tissues.

29 AURKA (Aurora kinase A) and AURKB (Aurora kinase B) belong to the Aurora

kinase subfamily and are conserved serine/threonine kinases (54). These two kinases are involved in mitosis and the maintenance of genomic integrity, and their deregulation can result in various abnormal mitotic events (54, 55). Many previous studies have found that the expression level and activity of these two kinases are upregulated in different tumors (56-58). Furthermore, Punt et al reported that the transcriptome and protein levels of Aurora kinase have significant negative correlations with immunocyte infiltration, the immunotherapeutic response and survival in patients with melanoma, and Aurora kinases were identified to be involved in the resistance of melanoma cells to T-cell-mediated cytotoxicity (59). In our study, AURKA and AURKB were significantly negatively correlated with mast cell infiltration and KITLG expression levels in the analysis of multiple databases. Therefore, we believe that AURKA and AURKB also inhibit the infiltration of mast cells in lung adenocarcinoma tissues through negatively regulating KITLG.

PLK1 (Polo-like kinase 1) belongs to the Polo-like kinase family, which plays an irreplaceable role in cell cycle regulation (60). Many studies have confirmed that the expression of PLK1 is elevated in various tumors, and its high expression level is associated with the poor prognosis of patients (61-63). By analyzing 33 cancer types from the TCGA database, Li et al reported that the high expression level of PLK1 always leads to lower immune activities, and the improvement of antitumor immune activities may increase the sensitivity of tumor cells to PLK1 inhibitors (64). In our study, PLK1 also showed strong negative correlations with mast cell infiltration levels and expression levels of KITLG, indicating that PLK1 may be involved in the negative regulation of KITLG expression and mast cell infiltration in lung adenocarcinoma tissues.

Enhancer of zeste homolog 2 (EZH2) belongs to the family of polycomb group genes (PcGs) and plays a vital role in transcriptional regulation (65). Previous studies have confirmed that EZH2 plays a crucial role in cell cycle regulation, apoptosis and other processes and is closely associated with the initiation, progression and metastasis of many tumors (66). Additionally, EZH2 plays an important role in the maintenance

1 of immune homeostasis by regulating various immunocytes (67). However, the
2 regulation of mast cells by EZH2 in the tumor microenvironment has not yet been
3 reported. In our study, the EZH2 and mast cell infiltration levels were significantly
4 negatively correlated in lung adenocarcinoma tissues. In the analysis of human normal
5 tissue transcriptome data (sigmoid colon, transverse colon, and terminal ileum), EZH2
6 and KITLG were also significantly negatively correlated. Therefore, we speculate that
7 EZH2 also regulates the transcription of KITLG, inhibiting the infiltration of mast cells
8 in early-stage lung adenocarcinoma tissues.

CONCLUSION

In our study, we described the protective effect of mast cells on patients with early-stage lung adenocarcinoma (stage I and stage II). In these patients, the mast cell infiltration levels are related to the mutation status of various driver/suppressor genes (TP53, STK11 and EGFR) and may play a certain role in predicting the efficacy of anti-PD-1/PD-L1 therapy. Additionally, by studying upstream regulators of mast cells, we found that CAMK2D, AURKA, AURKB, PLK1 and EZH2 are significantly associated with the mast cell infiltration levels and expression level of its key growth factor KITLG. These 5 upstream regulators are also closely related to the survival of patients with early-stage lung adenocarcinoma. Therefore, we speculate that these upstream regulators promote or inhibit the infiltration of mast cells in early-stage lung adenocarcinoma tissues by regulating KITLG. These research results will help elucidate the immune microenvironment of lung adenocarcinoma and the personalized immunotherapy of related patients.

Fig. 1 (A) Immunocytes infiltration proportions of 309 early-stage lung adenocarcinoma samples from GSE72094. (B) Mast cell infiltration proportions between stage I and stage II group of samples from GSE72094 (No statistical difference). (C) Kaplan–Meier curve of patients from GSE72094 when using the upper quartile of mast cell infiltration proportions as the cut-off value. (D) Sankey plot depicting the relationship across the tumor stage, TP53 mutation status and EGFR mutation status of patients stratified by median value of mast cell infiltration levels.

Fig. 2 Weighted gene co-expression network analysis (WGCNA) of early-stage lung adenocarcinoma samples from GSE72094. (A) Correlations between module eigengenes (MEs) and mast cell infiltration proportions and mutation status of key genes (TP53, STK11 and EGFR). (B) Gene significance (GS) of the genes contained in the brown module versus module membership (mast cell infiltration proportions). (C) Gene significance (GS) of the genes contained in the red module versus module membership (mast cell infiltration proportions).

Fig. 3 Upstream regulators of module eigengenes (MEs) and correlations across mast cells, KITLG, KIT and upstream regulators in GSE72094. (A) Upstream transcription factors (TFs, red hexagon nodes) and kinases (blue hexagon nodes) of the genes contained in the brown module (grey nodes) inferred from the Expression2Kinases (X2K) analysis. (B) Correlations across TFs (inferred from brown module), KITLG,

1 KIT and mast cell infiltration proportions. (C) Correlations across kinases (inferred
2 from brown module), KITLG, KIT and mast cell infiltration proportions. (D) Upstream
3 transcription factors (TFs, red hexagon nodes) and kinases (blue hexagon nodes) of the
4 genes contained in the red module (grey nodes) inferred from the Expression2Kinases
5 (X2K) analysis. (E) Correlations across TFs (inferred from red module), KITLG, KIT
6 and mast cell infiltration proportions. (F) Correlations across kinases (inferred from red
7 module), KITLG, KIT and mast cell infiltration proportions.

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

25

26

27

28

29

Fig. 4 Survival and correlation analysis of data from TCGA. (A) Kaplan–Meier curve when using the median of mast cell infiltration proportions as the cut-off value. (B) Sankey plot depicting the relationship across T stage, N stage and survival status of patients stratified by median value of mast cell infiltration levels. (C) Correlations across TFs (inferred from brown module), KITLG, KIT and mast cell infiltration proportions. (D) Correlations across kinases (inferred from brown module), KITLG, KIT and mast cell infiltration proportions. (E) Correlations across TFs (inferred from red module), KITLG, KIT and mast cell infiltration proportions. (F) Correlations across kinases (inferred from red module), KITLG, KIT and mast cell infiltration proportions.

Fig. 5 Correlations between KITLG and upstream regulators in normal tissues from GEPIA. GEPIA, Gene Expression Profiling Interactive Analysis. R, R value of Spearman's correlation.

Fig. 6 The potential interactions of tumor cells, mast cells, KITLG and related upstream regulators in early-stage LUAD tissues.

Fig. 7 Kaplan–Meier curve of KITLG, CAMK2D, AURKA, AURKB, PLK1 and EZH2 in GSE72094 when using the median of expression value as the cut-off value.

Table 1. Clinicopathologic characteristics of patients with early-stage LUAD in GSE72094

Characteristics	Mast cells infiltration		P-value ²
	High, N = 154 ¹	Low, N = 155 ¹	
Age	71 (64, 77)	70 (64, 76)	0.3
Gender			0.034
Female	92 (60%)	74 (48%)	
Male	62 (40%)	81 (52%)	
Smoking			0.6
Ever	114 (74%)	119 (77%)	
Missing	24 (16%)	25 (16%)	
Never	16 (10%)	11 (7.1%)	
Stage			0.6
I	124 (81%)	121 (78%)	
II	30 (19%)	34 (22%)	
EGFR status			0.032
Mut	24 (16%)	12 (7.7%)	
WT	130 (84%)	143 (92%)	
STK11 status			0.058
Mut	17 (11%)	29 (19%)	
WT	137 (89%)	126 (81%)	
TP53 status			<0.001
Mut	21 (14%)	52 (34%)	
WT	133 (86%)	103 (66%)	
KRAS status			0.6
Mut	54 (35%)	50 (32%)	
WT	100 (65%)	105 (68%)	

¹Median (IQR); n (%)

²Wilcoxon rank sum test; Pearson's Chi-squared test

Table 1. Clinicopathologic characteristics of patients with early-stage lung adenocarcinomas in GSE72094.

Table 2. Correlations of KITLG expression with upstream regulators in normal tissues

		KITLG				
		Bladder	Colon (Sigmoid)	Colon (Transverse)	Small Intestine (Terminal Ileum)	Stomach
CAMK2D	Cor	0.82	0.28	0.63	0.41	0.74
	P-value	0.0071*	0.00066*	0*	4.90E-05*	0*
AURKA	Cor	-0.25	-0.21	-0.43	-0.36	-0.19
	P-value	0.52	0.013*	1.00E-08*	0.00048*	0.011*
AURKB	Cor	-0.46	-0.25	-0.43	-0.42	-0.16
	P-value	0.21	0.003*	4.50E-09*	2.90E-05*	0.034*
PLK1	Cor	-0.44	-0.25	-0.4	-0.4	-0.2
	P-value	0.24	0.0029*	6.80E-08*	7.30E-05*	0.0066*
EZH2	Cor	-0.31	-0.34	-0.46	-0.52	-0.083
	P-value	0.42	4.80E-05*	4.40E-10*	7.90E-08*	0.27

Table 2. Correlation coefficient and P value of KITLG expression with upstream regulators in normal tissues utilizing the data from GTEx database. GTEx, Genotype-Tissue Expression. Cor, R value of Spearman's correlation.

1 **Abbreviations:** LUAD, lung adenocarcinoma; GEO, Gene Expression Omnibus;
2 TCGA, The Cancer Genome Atlas; GTEx, Genotype-Tissue Expression; WGCNA,
3 weighted gene coexpression network analysis; GDC, Genomic Data Commons; OS,
4 overall survival; MEs, module eigengenes; X2K, Expression2Kinases; TFs,
5 transcription factors; SCF, Stem cell factor; RTK, receptor tyrosine kinase; GEPIA,
6 Gene Expression Profiling Interactive Analysis; cor, R value of Spearman's correlation;
7 TME, tumor microenvironment; EGFR, epidermal growth factor receptor; NSCLC,
8 non-small cell lung cancer; EGFR-TKI, EGFR tyrosine kinase inhibitor; PD-L1,
9 programmed cell death ligand-1; PD-1, programmed cell death 1; CAMK2D,
10 calcium/calmodulin-dependent protein kinase II delta; AURKA, Aurora kinase A;
11 AURKB, Aurora kinase B; PLK1, Polo-like kinase 1; EZH2, Enhancer of zeste
12 homolog 2.

1 **Ethics approval and consent to participate**

2 Not applicable.

3 **Consent for publication**

4 Not applicable.

5 **Availability of data and materials**

6 The data that support the findings of this study are available in Gene Expression
7 Omnibus (GEO) and The Cancer Genome Atlas (TCGA). These data were derived from
8 the following resources available in the public domain:
9 <http://www.ncbi.nlm.nih.gov/geo/> and <https://portal.gdc.cancer.gov/>.

10 **Competing interests**

11 The authors declare that they have no competing interests.

12 **Funding**

13 None.

14 **Authors' contributions**

15 Peng Lin and Han Yang conceived this study; Bin Luo and Pengfei Duan retrieved and
16 disposed data from GEO database; Zhen Wang performed data analysis and prepared
17 the manuscript; Peng Lin and Han Yang revised the manuscript. All authors approved
18 the final version of the manuscript.

19 **Acknowledgements**

20 We thank the contributors to the profile of GSE72094 in the GEO datasets.

21 **Competing interests**

1 No conflict or financial interests.

2

3

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

25

26

27

28

29

REFERENCES

1. Bade BC, Dela Cruz CS. Lung Cancer 2020: Epidemiology, Etiology, and Prevention. *Clin Chest Med*. 2020;41(1):1-24.
2. Ma J, Ward EM, Smith R, Jemal A. Annual number of lung cancer deaths potentially avertable by screening in the United States. *Cancer*. 2013;119(7):1381-5.
3. Nanavaty P, Alvarez MS, Alberts WM. Lung cancer screening: advantages, controversies, and applications. *Cancer Control*. 2014;21(1):9-14.
4. Bach PB, Mirkin JN, Oliver TK, Azzoli CG, Berry DA, Brawley OW, et al. Benefits and harms of CT screening for lung cancer: a systematic review. *JAMA*. 2012;307(22):2418-29.
5. Chen YY, Huang TW, Tsai WC, Lin LF, Cheng JB, Chang H, et al. Risk factors of postoperative recurrences in patients with clinical stage I NSCLC. *World J Surg Oncol*. 2014;12:10.
6. Doroshow DB, Sanmamed MF, Hastings K, Politi K, Rimm DL, Chen L, et al. Immunotherapy in Non-Small Cell Lung Cancer: Facts and Hopes. *Clin Cancer Res*. 2019;25(15):4592-602.
7. Liu Y, Huang C, Liu H, Chen Y, Li S. Sublobectomy versus lobectomy for stage IA (T1a) non-small-cell lung cancer: a meta-analysis study. *World J Surg Oncol*. 2014;12:138.
8. Piliponsky AM, Acharya M, Shubin NJ. Mast Cells in Viral, Bacterial, and Fungal Infection Immunity. *Int J Mol Sci*. 2019;20(12).
9. Kumar V, Sharma A. Mast cells: emerging sentinel innate immune cells with diverse role in immunity. *Mol Immunol*. 2010;48(1-3):14-25.
10. Kirshenbaum AS, Goff JP, Semere T, Foster B, Scott LM, Metcalfe DD. Demonstration that human mast cells arise from a progenitor cell population that is CD34(+), c-kit(+), and expresses aminopeptidase N (CD13). *Blood*. 1999;94(7):2333-42.
11. Krystel-Whittemore M, Dileepan KN, Wood JG. Mast Cell: A Multi-Functional Master Cell. *Front Immunol*. 2015;6:620.
12. Dahlin JS, Hallgren J. Mast cell progenitors: origin, development and migration to tissues. *Mol Immunol*. 2015;63(1):9-17.
13. Theoharides TC, Valent P, Akin C. Mast Cells, Mastocytosis, and Related Disorders. *N Engl J Med*. 2015;373(2):163-72.
14. Virk H, Arthur G, Bradding P. Mast cells and their activation in lung disease. *Transl Res*. 2016;174:60-76.
15. Varricchi G, Galdiero MR, Loffredo S, Marone G, Iannone R, Marone G, et al. Are Mast Cells MASTers in Cancer? *Front Immunol*. 2017;8:424.
16. Wensman H, Kamgari N, Johansson A, Grujic M, Calounova G, Lundquist A, et al. Tumor-mast cell interactions: induction of pro-tumorigenic genes and anti-tumorigenic 4-1BB in MCs in response to Lewis Lung Carcinoma. *Mol Immunol*. 2012;50(4):210-9.
17. Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, et al. NCBI GEO: archive for functional genomics data sets--update. *Nucleic Acids Res*. 2013;41(Database issue):D991-5.
18. Cancer Genome Atlas Research N, Weinstein JN, Collisson EA, Mills GB, Shaw KR, Ozenberger BA, et al. The Cancer Genome Atlas Pan-Cancer analysis project. *Nat Genet*. 2013;45(10):1113-20.
19. Schabath MB, Welsh EA, Fulp WJ, Chen L, Teer JK, Thompson ZJ, et al. Differential association of STK11 and TP53 with KRAS mutation-associated gene expression, proliferation and immune surveillance in lung adenocarcinoma. *Oncogene*. 2016;35(24):3209-16.
20. Zhang Z, Hernandez K, Savage J, Li S, Miller D, Agrawal S, et al. Uniform genomic data analysis in the NCI Genomic Data Commons. *Nat Commun*. 2021;12(1):1226.

21. Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, et al. Robust enumeration of cell subsets from tissue expression profiles. *Nat Methods*. 2015;12(5):453-7.
22. T T. A Package for Survival Analysis in R. (<https://CRANR-projectorg/package=survival>). 2020.
23. Kassambara A. ggpubr: 'ggplot2' Based Publication Ready Plots. (<https://CRANR-projectorg/package=ggpubr>). 2020.
24. Biecek AKaMKaP. survminer: Drawing Survival Curves using 'ggplot2'. (<https://CRANR-projectorg/package=survminer>). 2020.
25. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*. 2008;9:559.
26. Clarke DJB, Kuleshov MV, Schilder BM, Torre D, Duffy ME, Keenan AB, et al. eXpression2Kinases (X2K) Web: linking expression signatures to upstream cell signaling networks. *Nucleic Acids Res*. 2018;46(W1):W171-W9.
27. Berger SI, Posner JM, Ma'ayan A. Genes2Networks: connecting lists of gene symbols using mammalian protein interactions databases. *BMC Bioinformatics*. 2007;8:372.
28. Lachmann A, Ma'ayan A. KEA: kinase enrichment analysis. *Bioinformatics*. 2009;25(5):684-6.
29. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res*. 2003;13(11):2498-504.
30. Wen LP, Fahrni JA, Matsui S, Rosen GD. Airway epithelial cells produce stem cell factor. *Biochim Biophys Acta*. 1996;1314(3):183-6.
31. Valent P, Spanblochl E, Sperr WR, Sillaber C, Zsebo KM, Agis H, et al. Induction of differentiation of human mast cells from bone marrow and peripheral blood mononuclear cells by recombinant human stem cell factor/kit-ligand in long-term culture. *Blood*. 1992;80(9):2237-45.
32. Williams DE, Eisenman J, Baird A, Rauch C, Van Ness K, March CJ, et al. Identification of a ligand for the c-kit proto-oncogene. *Cell*. 1990;63(1):167-74.
33. Simko TWaV. R package "corrplot": Visualization of a Correlation Matrix. (<https://githubcom/taiyun/corrplot>). 2017.
34. Tang Z, Li C, Kang B, Gao G, Li C, Zhang Z. GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses. *Nucleic Acids Res*. 2017;45(W1):W98-W102.
35. Consortium GT. The Genotype-Tissue Expression (GTEx) project. *Nat Genet*. 2013;45(6):580-5.
36. Altorki NK, Markowitz GJ, Gao D, Port JL, Saxena A, Stiles B, et al. The lung microenvironment: an important regulator of tumour growth and metastasis. *Nat Rev Cancer*. 2019;19(1):9-31.
37. Oldford SA, Marshall JS. Mast cells as targets for immunotherapy of solid tumors. *Mol Immunol*. 2015;63(1):113-24.
38. Tomita M, Matsuzaki Y, Edagawa M, Shimizu T, Hara M, Onitsuka T. Distribution of Mast Cells in Mediastinal Lymph Nodes from Lung Cancer Patients. *World J Surg Oncol*. 2003;1(1):25.
39. Takanami I, Takeuchi K, Naruke M. Mast cell density is associated with angiogenesis and poor prognosis in pulmonary adenocarcinoma. *Cancer*. 2000;88(12):2686-92.
40. Rajput AB, Turbin DA, Cheang MC, Voduc DK, Leung S, Gelmon KA, et al. Stromal mast cells in invasive breast cancer are a marker of favourable prognosis: a study of 4,444 cases. *Breast Cancer Res Treat*. 2008;107(2):249-57.
41. Henderson WR, Chi EY, Jong EC, Klebanoff SJ. Mast cell-mediated tumor-cell cytotoxicity. Role of the peroxidase system. *J Exp Med*. 1981;153(3):520-33.

- 1 42. Bhowmick NA, Moses HL. Tumor-stroma interactions. *Curr Opin Genet Dev.* 2005;15(1):97-
2 101.
- 3 43. He S, Walls AF. Human mast cell chymase induces the accumulation of neutrophils,
4 eosinophils and other inflammatory cells in vivo. *Br J Pharmacol.* 1998;125(7):1491-500.
- 5 44. Bao X, Shi R, Zhao T, Wang Y. Mast cell-based molecular subtypes and signature associated
6 with clinical outcome in early-stage lung adenocarcinoma. *Mol Oncol.* 2020;14(5):917-32.
- 7 45. Aran D, Hu Z, Butte AJ. xCell: digitally portraying the tissue cellular heterogeneity landscape.
8 *Genome Biol.* 2017;18(1):220.
- 9 46. Bonanno L, Zulato E, Pavan A, Attili I, Pasello G, Conte P, et al. LKB1 and Tumor Metabolism:
10 The Interplay of Immune and Angiogenic Microenvironment in Lung Cancer. *Int J Mol Sci.*
11 2019;20(8).
- 12 47. Mantovani F, Collavin L, Del Sal G. Mutant p53 as a guardian of the cancer cell. *Cell Death*
13 *Differ.* 2019;26(2):199-212.
- 14 48. Devarakonda S, Morgensztern D, Govindan R. Genomic alterations in lung adenocarcinoma.
15 *Lancet Oncol.* 2015;16(7):e342-51.
- 16 49. Khalil FK, Altioek S. Advances in EGFR as a Predictive Marker in Lung Adenocarcinoma. *Cancer*
17 *Control.* 2015;22(2):193-9.
- 18 50. Biton J, Mansuet-Lupo A, Pecuchet N, Alifano M, Ouakrim H, Arrondeau J, et al. TP53, STK11,
19 and EGFR Mutations Predict Tumor Immune Profile and the Response to Anti-PD-1 in Lung
20 Adenocarcinoma. *Clin Cancer Res.* 2018;24(22):5710-23.
- 21 51. Brown GR, Hem V, Katz KS, Ovetsky M, Wallin C, Ermolaeva O, et al. Gene: a gene-centered
22 information resource at NCBI. *Nucleic Acids Res.* 2015;43(Database issue):D36-42.
- 23 52. Belinky F, Nativ N, Stelzer G, Zimmerman S, Iny Stein T, Safran M, et al. PathCards: multi-
24 source consolidation of human biological pathways. *Database (Oxford).* 2015;2015.
- 25 53. Mann-Chandler MN, Kashyap M, Wright HV, Norozian F, Barnstein BO, Gingras S, et al. IFN-
26 gamma induces apoptosis in developing mast cells. *J Immunol.* 2005;175(5):3000-5.
- 27 54. Fu J, Bian M, Jiang Q, Zhang C. Roles of Aurora kinases in mitosis and tumorigenesis. *Mol*
28 *Cancer Res.* 2007;5(1):1-10.
- 29 55. Marumoto T, Zhang D, Saya H. Aurora-A - a guardian of poles. *Nat Rev Cancer.* 2005;5(1):42-
30 50.
- 31 56. Tatsuka M, Katayama H, Ota T, Tanaka T, Odashima S, Suzuki F, et al. Multinuclearity and
32 increased ploidy caused by overexpression of the aurora- and lpl1-like midbody-associated
33 protein mitotic kinase in human cancer cells. *Cancer Res.* 1998;58(21):4811-6.
- 34 57. Liao Y, Liao Y, Li J, Li J, Fan Y, Xu B. Polymorphisms in AURKA and AURKB are associated with
35 the survival of triple-negative breast cancer patients treated with taxane-based adjuvant
36 chemotherapy. *Cancer Manag Res.* 2018;10:3801-8.
- 37 58. Katayama H, Ota T, Jisaki F, Ueda Y, Tanaka T, Odashima S, et al. Mitotic kinase expression
38 and colorectal cancer progression. *J Natl Cancer Inst.* 1999;91(13):1160-2.
- 39 59. Punt S, Malu S, McKenzie JA, Manrique SZ, Doorduijn EM, Mbofung RM, et al. Aurora kinase
40 inhibition sensitizes melanoma cells to T-cell-mediated cytotoxicity. *Cancer Immunol Immunother.*
41 2021;70(4):1101-13.
- 42 60. van de Weerd BC, Medema RH. Polo-like kinases: a team in control of the division. *Cell Cycle.*
43 2006;5(8):853-64.
- 44 61. Liu Z, Sun Q, Wang X. PLK1, A Potential Target for Cancer Therapy. *Transl Oncol.*

1 2017;10(1):22-32.

2 62. Degenhardt Y, Lampkin T. Targeting Polo-like kinase in cancer therapy. Clin Cancer Res.

3 2010;16(2):384-9.

4 63. Strebhardt K, Ullrich A. Targeting polo-like kinase 1 for cancer therapy. Nat Rev Cancer.

5 2006;6(4):321-30.

6 64. Li M, Liu Z, Wang X. Exploration of the Combination of PLK1 Inhibition with Immunotherapy

7 in Cancer Treatment. J Oncol. 2018;2018:3979527.

8 65. Pasini D, Di Croce L. Emerging roles for Polycomb proteins in cancer. Curr Opin Genet Dev.

9 2016;36:50-8.

10 66. Duan R, Du W, Guo W. EZH2: a novel target for cancer treatment. J Hematol Oncol.

11 2020;13(1):104.

12 67. Nutt SL, Keenan C, Chopin M, Allan RS. EZH2 function in immune cell development. Biol Chem.

13 2020;401(8):933-43.