DR. XUANWEN BAO (Orcid ID : 0000-0002-6232-3783)

MR. TIANYU ZHAO (Orcid ID : 0000-0002-2696-273X)

Received Date : 28-Nov-2019

Revised Date : 28-Jan-2020

Accepted Date : 12-Mar-2020

Article type : Research Article

Mast cell-based molecular subtypes and signature associated with clinical outcome in early-stage lung adenocarcinoma

Xuanwen Bao [1, 2*‡], Run Shi [3*], Tianyu Zhao [4, 5], Yanfang Wang [6‡]

1) Institute of Radiation Biology, Helmholtz Center Munich, German Research Center for Environmental Health, 85764 Neuherberg, Germany.
2) Technical University Munich (TUM), 80333 Munich, Germany.
3) Department of Radiation Oncology, University Hospital, Ludwig Maximilian University of Munich, Munich, Germany.
4) Institute and Clinic for Occupational, Social and Environmental Medicine, University Hospital, Ludwig Maximilian University of Munich; Comprehensive Pneumology Center (CPC) Munich, member DZL; German Center for Lung Research, Munich, Germany.
5) Institute of Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany.

6) Ludwig-Maximilians-Universität München (LMU), 80539 Munich, Germany.

‡ Corresponding author:

Xuanwen Bao: xuanwen.bao@tum.de

Yanfang Wang: yangfang.wang@cup.uni-muenchen.de

Xuanwen Bao and Run Shi contributed equally to the manuscript.

ORCID:

Xuanwen Bao: 0000-0002-6232-3783

Run Shi: 0000-0003-4667-4812

Tianyu Zhao: 0000-0002-2696-273X

Yanfang Wang: 0000-0002-7989-5323

Running title: Mast cells predict outcome in early-stage LUAD

Abbreviations:

Lung adenocarcinoma (LUAD)

The Cancer Genome Atlas (TCGA)

Weighted correlation network analysis (WGCNA)

Gene Expression Omnibus (GEO)

single-sample gene set enrichment analysis (ssGSEA)

Molecular Oncology (2020) © 2020 The Authors. Published by FEBS Press and John Wiley & Sons Ltd.

gene set enrichment analysis (GSEA)
hazard ratio (HR)
Gene Ontology (GO)
Differently expressed gene (DEG)
principal component analysis (PCA)

Abstract

Mast cells are a major component of the immune microenvironment in tumour tissues and modulate tumour progression by releasing pro-tumorigenic and anti-tumorigenic molecules. Regarding the impact of mast cells on the outcomes of patients with lung adenocarcinoma (LUAD) patient, several published studies have shown contradictory results. Here, we aimed at elucidating the role of mast cells in early-stage LUAD. We found that high mast cell abundance was correlated with prolonged survival in early-stage LUAD patients. The mast cell-related gene signature and gene mutation datasets were used to stratify early-stage LUAD patients into two molecular subtypes (subtype one and subtype two). The neural network-based framework constructed with the mast cell-related signature showed high accuracy in predicting response to immunotherapy. Importantly, the prognostic mast cell-related signature predicted the survival probability and the potential relationship between TP53 mutation, c-MYC activation and mast cell activities. The meta-analysis confirmed the prognostic value of the mast cell-related gene signature. In summary, this study might improve our understanding of the role of mast cells in early-stage LUAD and aid in the development of immunotherapy and personalized treatments for early-stage LUAD patients.

Keywords: early-stage lung adenocarcinoma (LUAD); mast cell; prognosis; immunotherapy

Introduction

Lung adenocarcinoma (LUAD) is one of the most complex and heterogeneous malignancies, both in molecular and phenotypic terms (Li et al., 2016; Mao et al., 2016; Wang et al., 2019a). The incidence of LUAD has been increasing in recent years. The treatment for early-stage LUAD includes operation, chemotherapy and radiotherapy (Besse et al., 2015). Additionally, immunotherapy serves a promising therapeutic strategy in many cancer types (Bao et al., 2019b; Couzin-Frankel, 2013; Purwar et al., 2012; Schumacher and Schreiber, 2015; Wasiuk et al., 2012) as well. However, there is still a long way to go for immunotherapy in LUAD. A clear understanding of the tumour immune microenvironment may aid in the development of immunotherapy for LUAD patients.

Mast cell is widely distributed in different tissues and is a major component of the immune microenvironment in tumour tissues. Mast cells modulate tumour initiation and progression through the secretion of pro-tumorigenic and anti-tumorigenic molecules (Varricchi et al., 2017). The controversial roles of mast cells result in conflicting effects among different tumour types (Alì et al., 2009; Carlini et al., 2010; Gounaris et al., 2007; Jeong et al., 2013; Nordlund and Askenase, 1983; Sinnamon et al., 2008; Welsh et al., 2005; Yang et al., 2011). Regarding the impact of mast cells on LUAD patient clinical outcomes, several contradictory results have been published (Carlini et al., 2010; Imada et al., 2000; Kurebayashi et al., 2016; Li et al., 2018; Nagata et al., 2003; Takanami et al., 2000). Although the results of these studies are quite different with respect to the prognostic values of mast cells, previous studies have shown that mast cell infiltration is more intensive in well-differentiated tumours and low-grade histologic subtypes than in poorly differentiated and high-grade subtypes (Carlini et al., 2010; Nagata et al., 2003). Thus, in this study, we focused on the effect of mast cells in early-stage LUAD patients. We analysed the potential role of mast cell, mast cell-related genes, and immunotherapy

outcomes in early-stage LUAD using bioinformatics models and machine learning methods.

Method

Data processing

The Cancer Genome Atlas (TCGA) transcriptome data, mutation data and clinical information were downloaded via the UCSC Xena browser (https://xenabrowser.net/). GSE11969, GSE13213, GSE29013, GSE30219, GSE31210, GSE37745, GSE42127, GSE50081, and GSE72094 were downloaded from the Gene Expression Omnibus (GEO) database (http://www.ncbi.nlm.nih.gov/geo/). The detailed TCGA clinical information is summarized in Table 1 and supplementary file 1.

Estimation of the abundance of immune cell populations and implementation of weighted correlation network analysis (WGCNA)

Transcriptome file of TCGA early-stage LUAD was applied on xCELL to estimate the abundance of different immune cell populations (Aran et al., 2017; Newman et al., 2015). WGCNA was accomplished with the R package “WGCNA” (Bao et al., 2019a; Langfelder and Horvath, 2008; Wang et al., 2019b). The expression profile of immune-related gene (from https://www.innatedb.com/redirect.do?go=resourcesGeneLists) was applied as the input of WGCNA. Gene significance quantified the association of individual genes with mast cell density, and module membership represented the correlation between module eigengenes and gene expression profiles. A power of $\beta = 3$ and a scale-free $R^2 = 0.95$ were set as soft-threshold parameters to ensure a signed scale-free co-expression gene network. A total of 6 non-grey modules were generated. Among these modules, the yellow module depicting the highest correlation (r=0.92, p=4.2e-115) was considered the most correlated with mast cell density. Survival analysis was performed using the R package “survival”. Cox regression analysis was used to determine the hazard ratio (HR). All genes in the yellow module were subjected to univariate Cox regression. The 110 genes that significantly associated with the survival of early-stage LUAD patients in the yellow module were identified as the mast cell-related gene signature. These identified genes

were applied on Gene Ontology (GO) analysis with the "clusterProfiler" package (Yu et al., 2012) to elucidate the potential mechanism behind the gene signature. R software (version: 3.5.3) was used for all the analyses in the manuscript.

Molecular subtype identification

The R package "CancerSubtypes" was applied to perform molecular subtype identification (Xu et al., 2017). Transcriptome profile and gene mutation datasets were used to perform cancer subtype analysis. The default parameters were used to perform the classification. The cluster number was selected as 2. Gene set enrichment analysis (GSEA) was performed with GSEA software from Broad Institute.

Differently expressed gene (DEG) analysis

The DEG analysis was performed with "Limma" package (Smyth, 2005). An empirical Bayesian method was applied to estimate the fold change between the molecular subtype one and two using moderated t-tests. The adjusted p-value for multiple testing was calculated using the Benjamini-Hochberg correction. The genes with an adjusted p-value less than 0.05 and absolute $\log_2$ (log to base two of) fold change greater than 1.5 were identified as DEGs between two molecular subtypes. GO analysis was performed based on the significant genes.

Prognostic gene signature-based risk score and ssGSEA implementation

The genes in the WGCNA yellow module were analysed with univariate Cox regression analysis. The comprehensive mast cell-related signature was calculated by PCA. The PCA-based risk score $MastCell_{pca}$ was derived from the first principal component of the 110 genes from mast cell-related gene signature. Let $E_{i,j}$ represent the $log2(RSEM+1)$ value of the key gene i in tumour sample j, and C_i represents the corresponding coefficient of the mast cell-related genes. The risk score $MastCell_{pca}$ was calculated as follows:

$$MastCell_{pca} = \begin{bmatrix} E_{11} & \cdots & E_{1j} \\ \vdots & \ddots & \vdots \\ E_{i1} & \cdots & E_{ij} \end{bmatrix} [C_1 \ldots C_i]^T$$

ssGSEA implementation and clinical response prediction

The enrichment scores of the hallmark genes were evaluated using single-sample GSEA (ssGSEA) with R package "GSVA" (Hänzelmann et al., 2013). The hallmark gene sets were obtained from MSigDB. Spearman's coefficient analysis was performed to analysing the correlation between prognostic gene signature-based risk score and each hallmark. The Tumor Immune Dysfunction and Exclusion (TIDE) algorithm was used to predict the clinical response to immune checkpoint blockade (Jiang et al., 2018).

Neural network construction

PyTorch was employed to construct the neural network to predict the immunotherapy response by the mast cell-related gene signature in Python (Version: 3.5) (Paszke et al., 2017). Stochastic gradient descent method and learning rate 0.001 were chosen for the optimizer of the model. Five layer was built with different input and output number. Batch normalization was performed in each layer. Dropout function (dropout rate: 0.2) was used in the training process but not in the testing process. Relu function was applied as the activate function. A logistic sigmoid function was used in the output layer. The python script is provided in supplementary file 2.

Random forest algorithm for feature importance ranking

A random forest algorithm was applied to find the most critical mutations associated with the mast cell signature-based risk score. Briefly, the gene mutation dataset (supplementary file 3) and mast cell signature-based risk score were applied to find the most important gene mutations associated with the mast cell signature-based risk score. First, the "ranger" package was used to find the best hyperparameter in the regression process (Wright and Ziegler, 2015). Then, the "randomforest" package was applied for the construction of the regression model (Liaw and Wiener, 2002). The R code for the analysis in the manuscript is in supplementary file 4.

Results

High mast cell abundance in early-stage LUAD benefits the survival of patients

The workflow of the manuscript is shown in Fig. 1A. To illustrate the correlation between mast cells and survival in early-stage LUAD patients, we first analysed the abundance of immune cell populations in early-stage LUAD tumours samples. We identified twenty-two immune cell populations and the correlations between these populations are shown in Fig. 1B. We found that high mast cell abundance benefited the survival of early-stage LUAD patients in the TCGA cohorts (Fig. 1C). To further confirm the association between mast cells and the survival of early-stage LUAD patients, we estimated the abundance of mast cells in two external cohorts (GSE31210 and GSE50081). The results showed that high mast cell abundance are associated with prolonged survival of early-stage LUAD patients, as we observed in the TCGA early-stage LUAD cohort (Fig. 1D and Fig. 1E).

Identification of a gene signature associated with mast cells

The immune-related genes were determined with WGCNA. Genes were clustered into seven modules (Fig. 2A). The correlation between the modules and mast cell abundance was calculated by Pearson's correlation coefficient (Fig. 2B). The yellow module showed the highest correlation coefficient with mast cells (Cor: 0.73). The plots of module membership and gene significance illustrated a significant correlation for each gene in the yellow module (cor: 0.92) (Fig. 2C). Then, each gene in the yellow module was analysed with a univariate Cox regression analysis. We identified 110 genes that were significantly associated with the survival of early-stage LUAD patients (Fig. 2D). The heatmap shows the expression level of the 110 genes (Fig. 2E). The 110 genes were defined as a mast cell-related gene signature (supplementary file 5) in early-stage LUAD patients. GO analysis revealed that cellular metabolic pathways, WNT signalling, antigen processing and presentation, and other enriched pathways were associated with the mast cell-related key genes (Fig. 2F).

Molecular subtype identification based on the mast cell-related gene signature in early-stage LUAD

As we observed, two expression patterns were identified in the expression profiles of

mast cell-related genes from expression heatmap of mast cell-related gene signature. We asked whether the mast cell-related gene signature could distinguish the molecular subtypes of early-stage LUAD. Using a combination of gene mutation datasets (genome characteristics) and the expression profiles of mast cell-related key gene signature (genetic characteristics), we performed molecular subtype identification on early-stage LUAD patients. Three methods were applied to show the classification effect of the molecular subtypes: (1) a clustering heatmap was generated to intuitively visualize the effect of sample clustering (Fig. 3A); (2) univariate Cox and Kaplan-Meier analyses were used to evaluate the significance of the difference in survival profiles between subtypes (HR=0.59) (Fig. 3B); and (3) the average silhouette width, a measure of cluster coherence, was calculated to appraise whether samples were more similar within or across subtypes (Fig. 3C). The results above indicated that mast cell-related key genes could stratify early-stage LUAD into two molecular subtypes (subtype one and subtype two) with distinct clinical and molecular characteristics. Tumors of molecular subtype two had greater average mast cell densities compared with tumors of molecular subtype one.

DEG analysis was performed to identify the DEGs between the subtype one and subtype two molecular subtypes. The heatmap shows the expression profile of the DEGs (adjusted p-value < 0.05 & $\log_2$ (FC) > 1.5) (Fig. 3D). Then, the DEGs were subjected to GO analysis (Fig. 3E). The results revealed enrichments in cell cycle-related terms. GSEA was performed on the subtype one and subtype two of early-stage LUAD. Upregulated pathways included pathways related to coagulation, inflammatory response and myogenesis in the subtype one (Fig. 3F). Downregulated pathways included pathways related to E2F targets, G2M checkpoints and MYC targets in the subtype one (Fig. 3G). The immune cell population distribution in the subtype one and subtype two further illustrated the different tumour immune microenvironments in the two molecular subtypes of early-stage LUAD (Fig. 3H). Among all immune cell populations, mast cells showed the most significant difference between the subtype one and subtype two (Fig. 3I).

Neural network-based model to identifying immunotherapy treatment outcomes

To further utilize the mast cell-related gene signature we identified, we built a neural network-based framework to predict which patient would respond to immunotherapy according to mast cell-related key genes. The detailed code is provided in supplementary file 2. Fig. 4A illustrates a diagram of the neural network. Briefly, the early-stage LUAD dataset was divided into training and testing datasets. We constructed the neural network with the mast cell-related gene signature by the training dataset. The test dataset was applied to evaluate the accuracy of the neural network. With the increased epoch number for training, the loss value of the model in the testing set decreased (Fig. 4B). The confusion matrix showed only one sample was recognized wrongly in the testing set (Fig. 4C). The receiver operating characteristic (ROC) curve illustrated a high accuracy rate with the area under the curve reaching 98.7% (Fig. 4D).

Mast cell-related signature predicts the prognosis and clinical outcome of early-stage LUAD patients

The mast cell-related gene signature was employed to calculate a prognostic risk score. The risk score $MastCell_{pca}$ was calculated for each patient using the PCA method. Fig. 5A shows the first principal component (PCA1) score for each key mast cell-related gene. $MastCell_{pca}$ was calculated with the expression level of each gene and the PCA1 score. The results showed a highly negative correlation between $MastCell_{pca}$ and mast cell abundance, which further confirmed the correlation between the mast cell-related key genes and mast cells (Fig. 5B). The Kaplan–Meier plot revealed that patients with a low-risk score had a better prognosis than patients with a high-risk score (Fig. 5C). ssGSEA results revealed a high association of DNA repair and the c-MYC pathway with the mast cell-related risk score (Fig. 5D). TP53 mutations can regulate the activation of c-MYC pathway (Frazier et al., 1998). Due to the high correlation between the c-MYC pathway and the mast cell-related risk score, we selected TP53 mutation as an example given its role in regulating the activation of c-MYC pathway (Frazier et al., 1998) and analysed the mast cell-related risk score in TP53-mutated and wild-type patients (Fig. 5E-5F). The results showed a high-risk score in the TP53-mutated patients.

Furthermore, patients with high mast cell abundance had a low mast cell-related risk score and responded to immunotherapy (Fig. 5G). In patients who received chemo(radio)therapy and molecular therapy, the patients with low mast cell-related risk scores had better survival outcomes than those with nigh mast cell-related risk scores (Fig. 5H-5I).

The association between mast cell-related signature and gene mutation in early-stage LUAD

The random forest algorithm was employed to determine the importance of gene mutations associated mast cell-related risk score (Fig. 6). The results revealed that TP53 and CSMD3 were the most important gene mutations associated with the mast cell-related risk score. The patients with TP53 mutations had significantly higher mast cell-related risk scores than the patients without TP53 mutations.

External validation and meta-analysis

Nine external cohorts were used to confirm the association between the mast cell-related gene signature and survival outcomes in early-stage LUAD patients. The detailed information for each cohort is shown in the Kaplan–Meier plot (Fig. 7A). A meta-analysis was performed with a random-effects model, and the results showed that patients with a high mast cell-related risk score had poor survival outcomes in the overall dataset (HR = 3.79) (Fig. 7B).

Discussion

Previous studies have investigated the relationship between immune cell populations and the clinical outcomes of cancer patients (Bao et al., 2019b; Bindea et al., 2013; Chung et al., 2017; Homma et al., 2014). The heterogeneity of immune cell populations in different cancer types leads to a complicated immune network in the tumour microenvironment and differentially influences tumour initiation and progression. As a major component of the immune microenvironment in tumour tissues, mast cells may play

a pro-tumorigenic or antitumorigenic role by releasing different mediators (Varricchi et al., 2017). For instance, angiogenic and lymphangiogenic factors secreted by mast cells promote tumour angiogenesis and lymphangiogenesis (Detoraki et al., 2010; Detoraki et al., 2009; Theoharides et al., 2010). Several matrix metalloproteinases released by mast cells regulate the digestion of tumour extracellular matrix and favour the distant metastasis of cancer cells (Baram et al., 2001). Specifically, activation of MYC triggers rapid recruitment of mast cells to the tumor site to promote tumor expansion in pancreatic cancer. MYC directly commandeers and instructs tissue remodeling, angiogenesis, and inflammation by activation of mast cells (Soucek et al., 2007). Mast cells release tryptase AB1 and Interleukin-1β, which in turn induced pleural vasculature leakiness and triggered NF-κB activation in pleural tumor cells, thereby fostering pleural fluid accumulation and tumor growth (Giannou et al., 2015). In contrast, mast cells can exhibit anti-tumour activity directly through tumour cell cytotoxicity mediated by TNF-α and ROS or indirectly through the release of Interleukin-9 and heparin and the stimulation of dendritic cell maturation (Varricchi et al., 2017). The complicated roles of mast cells allow them to play different functions in different cancer types and stages.

Regarding the impact of mast cells on LUAD patient outcomes, several contradictory studies have been published. One study revealed that mast cells correlated with angiogenesis and poor outcome in stage I LUAD (Baram et al., 2001). Another study has revealed KIT-competent mast cells fuel KRAS-mutant lung adenocarcinoma formation, growth, and metastasis by providing Interleukin-1β and are associated with LUAD progression (Lilis et al., 2019). However, one research indicated that only mast cells were found by univariate analysis to be associated with better prognosis in LUAD (Kurebayashi et al., 2016). Although the results of these studies are quite different with respect to the prognostic value of mast cells, previous studies have shown that mast cell infiltration is more intensive in low-grade histologic subtypes than in high-grade subtypes (Carlini et al., 2010). Understanding the potential mechanism and roles of mast cells in early-stage LUAD may be helpful for the development of immunotherapy. Thus, in this study, we

analysed the potential role and mast cell-related genes in early-stage LUAD. The abundance of mast cells was estimated in several cohorts. Cox regression was performed to identify the prognostic value of mast cells in early-stage LUAD. WGCNA was employed to identify the mast cell-related gene signature. Molecular subtypes (subtype one and subtype two) were identified according to the mast cell-related gene signature in a mutation dataset of early-stage LUAD. A neural network-based framework was constructed to predict the immunotherapy outcome of early-stage LUAD patients according to the mast cell-related gene signature. A mast cell-related risk score $MastCell_{pca}$ was calculated by the expression levels of mast cell-related gene signature using the PCA method. ssGSEA was performed to identify the potential molecular mechanism associated with the mast cell-related prognostic signature. The association between gene mutations and the risk scores was identified by a random forest algorithm. A meta-analysis was performed to validate the mast cell-related signature in external cohorts.

In our analysis, we revealed that a high abundance of mast cells was associated with prolonged survival in early-stage LUAD patients. Two external cohorts confirmed this conclusion. The differences and controversial conclusions in different studies (Baram et al., 2001; Kurebayashi et al., 2016; Lilis et al., 2019) may be due to the mixture of activated and resting mast cells. The function of activated mast cells may be masked by the resting mast cells. Therefore, it is essential to analyse the activated and resting mast cells separately. In an alternative way, we employed the following workflow to identify the potential mechanisms and genes associated with mast cells.

First, the immune-related genes were clustered into several modules by unsupervised clustering. The yellow module was identified as the most important module correlated with mast cells according to Pearson's correlation coefficient. The mast cell-related gene signature was obtained from the yellow module. The genes in the mast cell-related gene signature were highly asscoaited with the mast cell density in early-stage LUAD. According to the mast cell-related gene signature and genome characteristics, the

early-stage LUAD tissues were stratified into two molecular subtypes (subtype one and subtype two). Interestingly, the GO analysis and GSEA both indicated enrichments in cell cycle and c-MYC related pathways in the subtype two. Thus, we concluded the potential involvement of mast cells in the c-MYC pathway in early-stage LUAD. One previous study has demonstrated the important roles of mast cells in MYC activation and the potential tumor expansion promoted by mast cells in pancreatic cancer. MYC is a highly pleiotropic transcription factor whose aberrant activation links tightly with tumor progression, including both cell-intrinsic proliferation and extracellular microenvironment alterations such as tissue remodelling, angiogenesis and invison (Gabay et al., 2014). Aberrant MYC activities induces the dysregulated expression of a chemokine-encoding gene cluster, therefore chemoattracting mast cells into the islets of pancreatic cancer (Soucek et al., 2007). In consistent with the pancreatic cancer study, the transcriptomic and downstream analysis underscore the importance of mast cell in MYC activation in early-stage LUAD. Moreover, the differences in mast cells between the two subtypes were the most significant of all the immune cell populations studied, which confirmed the relationship between the mast cell-related gene signature and mast cell abundance. The mast cell-related gene signature may represent targets for further study to aid in the understanding of the mechanism of mast cells in early-stage LUAD.

To further utilize the mast cell-related gene signature, we built a neural network-based framework to predict response to immunotherapy. The confusion matrix and ROC plot confirmed the accuracy of the network's prediction capability. Hence, we were able to apply the expression profile of the mast cell-related genes to predict the response to immunotherapy using the neural network framework.

In the next step, we calculated the risk score $MastCell_{pca}$ according to the expression level of the gene signature for each patient. ssGSEA revealed a significant correlation between DNA repair, the c-MYC pathway, and the signature-based risk score. The ssGSEA results further confirmed the results from the canonical GSEA of the molecular subtypes. c-MYC stimulates the expression of target genes that play important roles in cell

proliferation, growth arrest and apoptosis in lung cancer cells (Dang et al., 2006; Tong et al., 2004). Additionally, we further identified TP53 as the most critical mutation associated with the mast cell-related signature. Dysregulation of the c-MYC pathway induces the expression of endogenous TP53. As a cellular gatekeeper, TP53 plays crucial role in cell cycle arrest and apoptosis (Mogi and Kuwano, 2011). The close link between the mast cell-related signature, the c-MYC pathway and TP53 mutation in our analysis may highlight the roles of mast cells in early-stage LUAD. However, as we suggested, analysing the activated and resting mast cells separately would be a promising way for understanding the molecular mechanism of mast cells in early-stage LUAD. The mast cell-related gene signature we obtained may therefore prove useful information for further study in this field.

The mast cell-related signature also served as a promising marker to predict the survival of early-stage LUAD patients. We performed a meta-analysis by combining nine cohorts. The results revealed in both each cohort and the meta-analysis that the mast cell-related signature stratified the survival of patients with high and low signature-based risk scores. The results above also confirmed the pivotal roles of mast cells in early-stage LUAD.

A problem with the mast cell-related signature of early-stage LUAD as shown is that only in-silico analysis is performed. Experimental studies are required to further elucidate the biological functions underlying the mast cell-related signature in early-stage LUAD. Large, well-designed prospective population-based studies should be conducted to investigate the complex role of mass cell and testify our results on mast cell-related signature.

Conclusion

In this study, we depicted the correlation between mast cell populations and prognosis in early-stage LUAD patients. A mast cell-related gene signature was identified. A novel molecular subtype classification and a mast cell-related gene signature-based neural network were built to help understanding of mast cell activities in early-stage LUAD and

aid in the development of immunotherapy for early-stage LUAD patients. Potential pathways associated with the mast cell-related gene signature provide new directions for determining novel mechanisms in mast cells in early-stage LUAD. The results above may facilitate personalized medicine for early-stage LUAD patients.

Declarations

Ethics approval and consent to participate

Not applicable.

Conflict of interest statement

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

Data accessibility

The datasets supporting the conclusions of this article are available in the Xena browser (https://xenabrowser.net/) repository.

Funding

We greatly thank the China Scholarship Council (CSC) for supporting the research and work of Xuanwen Bao (No. 201608210186), Tianyu Zhao (No. 201708120056) and Run Shi (No. 201708320347).

Author Contributions

XW. B., YF. W. and R. S. conceived and designed the experiments. XW. B. and R. S. performed the analysis. XW. B. and YF. W. wrote the paper. XW. B., TY. Z., R. S. and YF. W. reviewed the draft. All authors read and approved the final manuscript.

Acknowledgements

We would like to thank Haowen Deng for helpful discussions and suggestions.

Table 1:

Patient information

Variable	Number
Gender (female/male)	166/142
TNM Stage (stage I/stage II)	212/96
Lymph node metastasis (positive/negative)	61/247
Age (>60/<=60/missing)	221/78/9
KRAS Status (WT/MUT)	210/98
EGFR Status (WT/MUT)	271/37
Smoking (no/yes/missing)	15/93/200

Supplementary file 1: Patient clinical information

Supplementary file 2: The python script for building the neural network-based framework.

Supplementary file 3: The gene mutation dataset.

Supplementary file 4: The R script.

Supplementary file 5: The gene list of the mast cell-related gene signature.

References

Alì, G., Boldrini, L., Lucchi, M., Mussi, A., Corsi, V., Fontanini, G., 2009. Tryptase mast cells in malignant pleural mesothelioma as an independent favorable prognostic factor. Journal of Thoracic Oncology 4, 348-354.

Aran, D., Hu, Z., Butte, A.J., 2017. xCell: digitally portraying the tissue cellular heterogeneity landscape. Genome biology 18, 220.

Bao, M., Shi, R., Zhang, K., Zhao, Y., Wang, Y., Bao, X., 2019a. Development of a membrane lipid metabolism–based signature to predict overall survival for personalized medicine in ccRCC patients. EPMA Journal.

Bao, X., Shi, R., Zhang, K., Xin, S., Li, X., Zhao, Y., Wang, Y., 2019b. Immune Landscape of Invasive Ductal Carcinoma

Tumor Microenvironment Identifies a Prognostic and Immunotherapeutically Relevant Gene Signature. Front. Oncol 9, 903.

Baram, D., Vaday, G.G., Salamon, P., Drucker, I., Hershkoviz, R., Mekori, Y.A., 2001. Human mast cells release metalloproteinase-9 on contact with activated T cells: juxtacrine regulation by TNF-α. The Journal of Immunology 167, 4008-4016.

Besse, B., Johnson, M., Janne, P., Garassino, M., Eberhardt, W., Peters, S., Toh, C., Kurata, T., Li, Z., Kowanetz, M., 2015. 16LBA Phase II, single-arm trial (BIRCH) of atezolizumab as first-line or subsequent therapy for locally advanced or metastatic PD-L1-selected non-small cell lung cancer (NSCLC). European Journal of Cancer 51, S717-S718.

Bindea, G., Mlecnik, B., Tosolini, M., Kirilovsky, A., Waldner, M., Obenauf, A.C., Angell, H., Fredriksen, T., Lafontaine, L., Berger, A., 2013. Spatiotemporal dynamics of intratumoral immune cells reveal the immune landscape in human cancer. Immunity 39, 782-795.

Carlini, M.J., Dalurzo, M.C.L., Lastiri, J.M., Smith, D.E., Vasallo, B.C., Puricelli, L.I., de Cidre, L.S.L., 2010. Mast cell phenotypes and microvessels in non–small cell lung cancer and its prognostic significance. Human pathology 41, 697-705.

Chung, W., Eum, H.H., Lee, H.-O., Lee, K.-M., Lee, H.-B., Kim, K.-T., Ryu, H.S., Kim, S., Lee, J.E., Park, Y.H., 2017. Single-cell RNA-seq enables comprehensive tumour and immune cell profiling in primary breast cancer. Nature communications 8, 15081.

Couzin-Frankel, J., 2013. Cancer immunotherapy. American Association for the Advancement of Science.

Dang, C.V., O'Donnell, K.A., Zeller, K.I., Nguyen, T., Osthus, R.C., Li, F., 2006. The c-Myc target gene network, Seminars in cancer biology. Elsevier, pp. 253-264.

Detoraki, A., Granata, F., Staibano, S., Rossi, F., Marone, G., Genovese, A., 2010. Angiogenesis and lymphangiogenesis in bronchial asthma. Allergy 65, 946-958.

Detoraki, A., Staiano, R.I., Granata, F., Giannattasio, G., Prevete, N., de Paulis, A., Ribatti, D., Genovese, A., Triggiani, M., Marone, G., 2009. Vascular endothelial growth factors synthesized by human lung mast cells exert angiogenic effects. Journal of Allergy and Clinical Immunology 123, 1142-1149. e1145.

Frazier, M.W., He, X., Wang, J., Gu, Z., Cleveland, J.L., Zambetti, G.P., 1998. Activation of c-myc gene expression by tumor-derived p53 mutants requires a discrete C-terminal domain. Molecular and cellular biology 18, 3735-3743.

Gabay, M., Li, Y., Felsher, D.W., 2014. MYC activation is a hallmark of cancer initiation and maintenance. Cold Spring Harbor perspectives in medicine 4, a014241.

Giannou, A.D., Marazioti, A., Spella, M., Kanellakis, N.I., Apostolopoulou, H., Psallidas, I., Prijovich, Z.M., Vreka, M., Zazara, D.E., Lilis, I., 2015. Mast cells mediate malignant pleural effusion formation. The Journal of clinical investigation 125, 2317-2334.

Gounaris, E., Erdman, S.E., Restaino, C., Gurish, M.F., Friend, D.S., Gounari, F., Lee, D.M., Zhang, G., Glickman, J.N., Shin, K., 2007. Mast cells are an essential hematopoietic component for polyp development. Proceedings of the National Academy of Sciences 104, 19977-19982.

Hänzelmann, S., Castelo, R., Guinney, J., 2013. GSVA: gene set variation analysis for microarray and RNA-seq data. BMC bioinformatics 14, 7.

Homma, Y., Taniguchi, K., Nakazawa, M., Matsuyama, R., Mori, R., Takeda, K., Ichikawa, Y., Tanaka, K., Endo, I., 2014. Changes in the immune cell population and cell proliferation in peripheral blood after gemcitabine-based chemotherapy for pancreatic cancer. Clinical and Translational Oncology 16, 330-335.

Imada, A., Shijubo, N., Kojima, H., Abe, S., 2000. Mast cells correlate with angiogenesis and poor outcome in stage I lung adenocarcinoma. European Respiratory Journal 15, 1087-1093.

Jeong, H.J., Oh, H.A., Nam, S.Y., Han, N.R., Kim, Y.S., Kim, J.H., Lee, S.J., Kim, M.H., Moon, P.D., Kim, H.M., 2013. The critical role of mast cell‐derived hypoxia‐inducible factor‐1α in human and mice melanoma growth. International journal of cancer 132, 2492-2501.

Jiang, P., Gu, S., Pan, D., Fu, J., Sahu, A., Hu, X., Li, Z., Traugh, N., Bu, X., Li, B., 2018. Signatures of T cell dysfunction and exclusion predict cancer immunotherapy response. Nature medicine, 1.

Kurebayashi, Y., Emoto, K., Hayashi, Y., Kamiyama, I., Ohtsuka, T., Asamura, H., Sakamoto, M., 2016. Comprehensive immune profiling of lung adenocarcinomas reveals four immunosubtypes with plasma cell subtype a negative indicator. Cancer immunology research 4, 234-247.

Langfelder, P., Horvath, S., 2008. WGCNA: an R package for weighted correlation network analysis. BMC bioinformatics 9, 559.

Li, S., Choi, Y.-L., Gong, Z., Liu, X., Lira, M., Kan, Z., Oh, E., Wang, J., Ting, J.C., Ye, X., 2016. Comprehensive characterization of oncogenic drivers in Asian lung adenocarcinoma. Journal of Thoracic Oncology 11, 2129-2140.

Li, X., Li, J., Wu, P., Zhou, L., Lu, B., Ying, K., Chen, E., Lu, Y., Liu, P., 2018. Smoker and non-smoker lung

adenocarcinoma is characterized by distinct tumor immune microenvironments. Oncoimmunology 7, e1494677.

Liaw, A., Wiener, M., 2002. Classification and regression by randomForest. R news 2, 18-22.

Lilis, I., Ntaliarda, G., Papaleonidopoulos, V., Giotopoulou, G.A., Oplopoiou, M., Marazioti, A., Spella, M., Marwitz, S., Goldmann, T., Bravou, V., 2019. Interleukin-1β provided by KIT-competent mast cells is required for KRAS-mutant lung adenocarcinoma. OncoImmunology 8, 1593802.

Mao, Y., Yang, D., He, J., Krasna, M.J., 2016. Epidemiology of lung cancer. Surgical Oncology Clinics 25, 439-445.

Mogi, A., Kuwano, H., 2011. TP53 mutations in nonsmall cell lung cancer. BioMed Research International 2011.

Nagata, M., Shijubo, N., Walls, A.F., Ichimiya, S., Abe, S., Sato, N., 2003. Chymase-positive mast cells in small sized adenocarcinoma of the lung. Virchows Archiv 443, 565-573.

Newman, A.M., Liu, C.L., Green, M.R., Gentles, A.J., Feng, W., Xu, Y., Hoang, C.D., Diehn, M., Alizadeh, A.A., 2015. Robust enumeration of cell subsets from tissue expression profiles. Nature methods 12, 453.

Nordlund, J.J., Askenase, P.W., 1983. The effect of histamine, antihistamines, and a mast cell stabilizer on the growth of cloudman melanoma cells in DBA/2 mice. Journal of Investigative Dermatology 81, 28-31.

Paszke, A., Gross, S., Chintala, S., Chanan, G., Yang, E., DeVito, Z., Lin, Z., Desmaison, A., Antiga, L., Lerer, A., 2017. Automatic differentiation in pytorch.

Purwar, R., Schlapbach, C., Xiao, S., Kang, H.S., Elyaman, W., Jiang, X., Jetten, A.M., Khoury, S.J., Fuhlbrigge, R.C., Kuchroo, V.K., 2012. Robust tumor immunity to melanoma mediated by interleukin-9–producing T cells. Nature medicine 18, 1248.

Schumacher, T.N., Schreiber, R.D., 2015. Neoantigens in cancer immunotherapy. Science 348, 69-74.

Sinnamon, M.J., Carter, K.J., Sims, L.P., LaFleur, B., Fingleton, B., Matrisian, L.M., 2008. A protective role of mast cells in intestinal tumorigenesis. Carcinogenesis 29, 880-886.

Smyth, G.K., 2005. Limma: linear models for microarray data, Bioinformatics and computational biology solutions using R and Bioconductor. Springer, pp. 397-420.

Soucek, L., Lawlor, E.R., Soto, D., Shchors, K., Swigart, L.B., Evan, G.I., 2007. Mast cells are required for angiogenesis and macroscopic expansion of Myc-induced pancreatic islet tumors. Nature medicine 13, 1211.

Takanami, I., Takeuchi, K., Naruke, M., 2000. Mast cell density is associated with angiogenesis and poor prognosis in pulmonary adenocarcinoma. Cancer 88, 2686-2692.

Theoharides, T.C., Zhang, B., Kempuraj, D., Tagen, M., Vasiadi, M., Angelidou, A., Alysandratos, K.-D., Kalogeromitros,

D., Asadi, S., Stavrianeas, N., 2010. IL-33 augments substance P–induced VEGF secretion from human mast cells and is increased in psoriatic skin. Proceedings of the National Academy of Sciences 107, 4448-4453.

Tong, X., O'Kelly, J., Xie, D., Mori, A., Lemp, N., McKenna, R., Miller, C.W., Koeffler, H.P., 2004. Cyr61 suppresses the growth of non-small-cell lung cancer cells via the β-catenin–c-myc–p53 pathway. Oncogene 23, 4847.

Varricchi, G., Galdiero, M.R., Loffredo, S., Marone, G., Iannone, R., Marone, G., Granata, F., 2017. Are mast cells MASTers in cancer? Frontiers in immunology 8, 424.

Wang, Y., Deng, H., Xin, S., Zhang, K., Shi, R., Bao, X., 2019a. Prognostic and predictive value of three DNA methylation signatures in lung adenocarcinoma. Frontiers in genetics 10, 349.

Wang, Y., Zhang, Q., Gao, Z., Xin, S., Zhao, Y., Zhang, K., Shi, R., Bao, X., 2019b. A novel 4-gene signature for overall survival prediction in lung adenocarcinoma patients with lymph node metastasis. Cancer Cell International 19, 100.

Wasiuk, A., Dalton, D.K., Schpero, W.L., Stan, R.V., Conejo-Garcia, J.R., Noelle, R.J., 2012. Mast cells impair the development of protective anti-tumor immunity. Cancer immunology, immunotherapy 61, 2273-2282.

Welsh, T.J., Green, R.H., Richardson, D., Waller, D.A., O'Byrne, K.J., Bradding, P., 2005. Macrophage and mast-cell invasion of tumor cell islets confers a marked survival advantage in non–small-cell lung cancer. Journal of clinical oncology 23, 8959-8967.

Wright, M.N., Ziegler, A., 2015. ranger: A fast implementation of random forests for high dimensional data in C++ and R. arXiv preprint arXiv:1508.04409.

Xu, T., Le, T.D., Liu, L., Su, N., Wang, R., Sun, B., Colaprico, A., Bontempi, G., Li, J., 2017. CancerSubtypes: an R/Bioconductor package for molecular cancer subtype identification, validation and visualization. Bioinformatics 33, 3131-3133.

Yang, X.D., Ai, W., Asfaha, S., Bhagat, G., Friedman, R.A., Jin, G., Park, H., Shykind, B., Diacovo, T.G., Falus, A., 2011. Histamine deficiency promotes inflammation-associated carcinogenesis through reduced myeloid maturation and accumulation of CD11b+ Ly6G+ immature myeloid cells. Nature medicine 17, 87.

Yu, G., Wang, L.-G., Han, Y., He, Q.-Y., 2012. clusterProfiler: an R package for comparing biological themes among gene clusters. Omics: a journal of integrative biology 16, 284-287.

Figure legends

Fig. 1 The association between mast cell abundance and clinical outcomes in

early-stage LUAD patients. (A) Schematic diagram of the study design. (B) The correlation among immune cell populations. (C-E) Kaplan-Meier curves for the OS of early-stage LUAD patients showed that the patients with high mast cell abundance had a favourable outcome compared with the patients with low mast cell abundance in the TCGA, GSE31210 and GSE50081 cohorts.

Fig. 2 Mast cell-related gene signature identification. (A) WGCNA was performed to identify seven modules by unsupervised clustering. (B) A total of six modules (non-grey) were identified. The yellow module had the highest correlation ($r=0.73$, $p=8e^{-53}$) and was considered the most correlated with mast cells. (C) The gene significance and module membership of the genes in the yellow module exhibited a high correlation. (D) A total of 110 mast cell-related genes were identified among the hub genes extracted from the yellow module. (E) The expression profile of the 110 mast cell-related genes. (F) GO analysis was performed based on the 110 mast cell-related genes.

Fig. 3 Molecular subtype identification according to the mast cell-related gene signature. (A) Clustering heatmap for intuitively visualizing the effect of sample clustering. (B) Univariate Cox analysis and Kaplan-Meier curves were used to evaluate the survival difference between the two molecular subtypes. (C) Average silhouette width between the two molecular subtypes. (D) The DEGs between 2 molecular subtypes. (E) GO analysis. (F) Upregulated hallmarks in the GSEA. (G) Downregulated hallmarks in the GSEA. (H) The immune cell population distribution in the subtype one and subtype two. (I) The difference in immune cell population scores and the significances between the subtype one and subtype two.

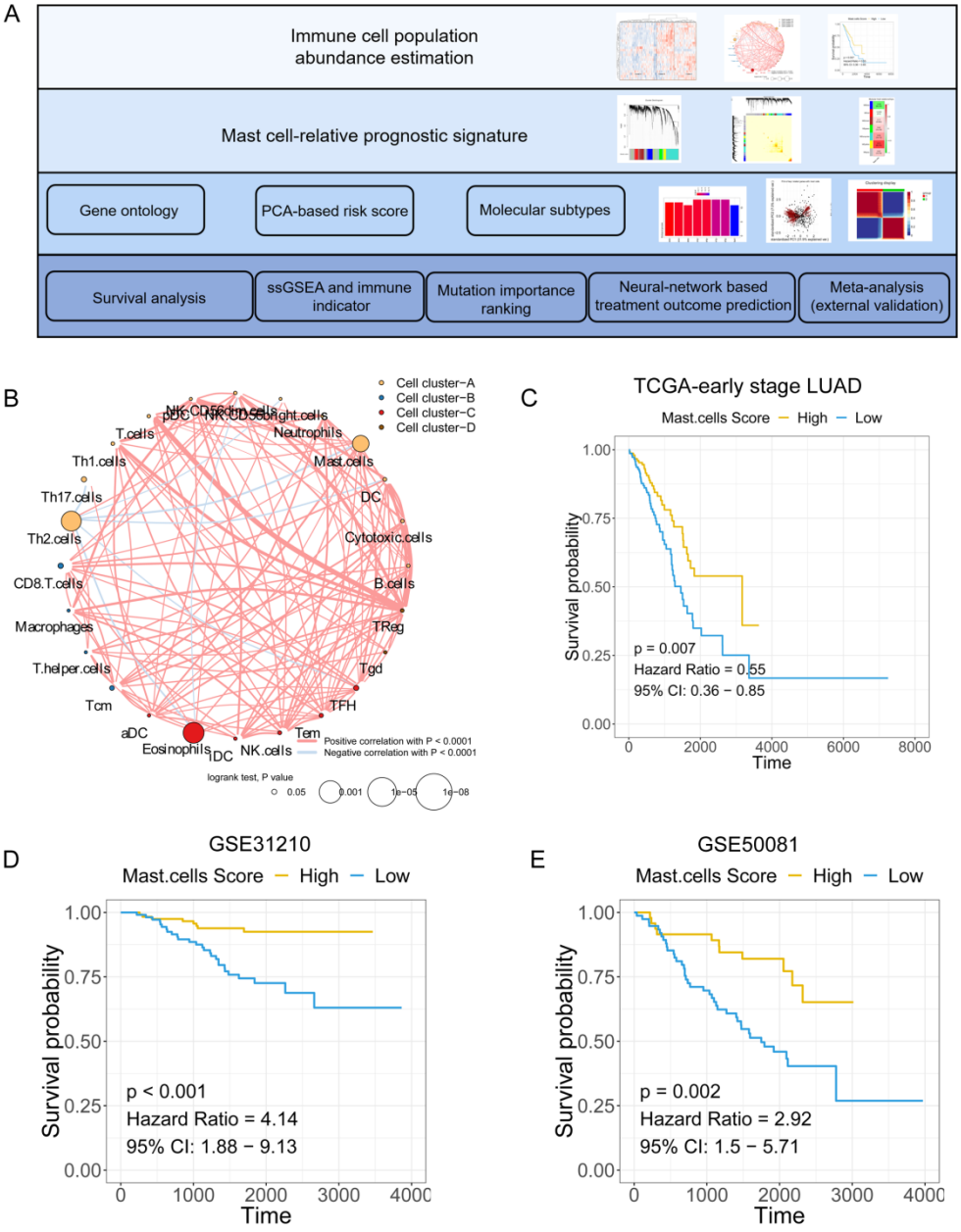
Fig. 4 Neural network-based framework construction with the mast cell-related gene signature. (A) Schematic diagram of the neural network. (B) The loss value in each epoch during training process in the validation cohort. (C) The confusion matrix in the testing cohort validated the accuracy of the network's prediction capacity. (D) The ROC plot in the testing dataset validated the accuracy of the network's prediction capacity.

Fig. 5 Mast cell-related signature-based risk score calculation and the potential

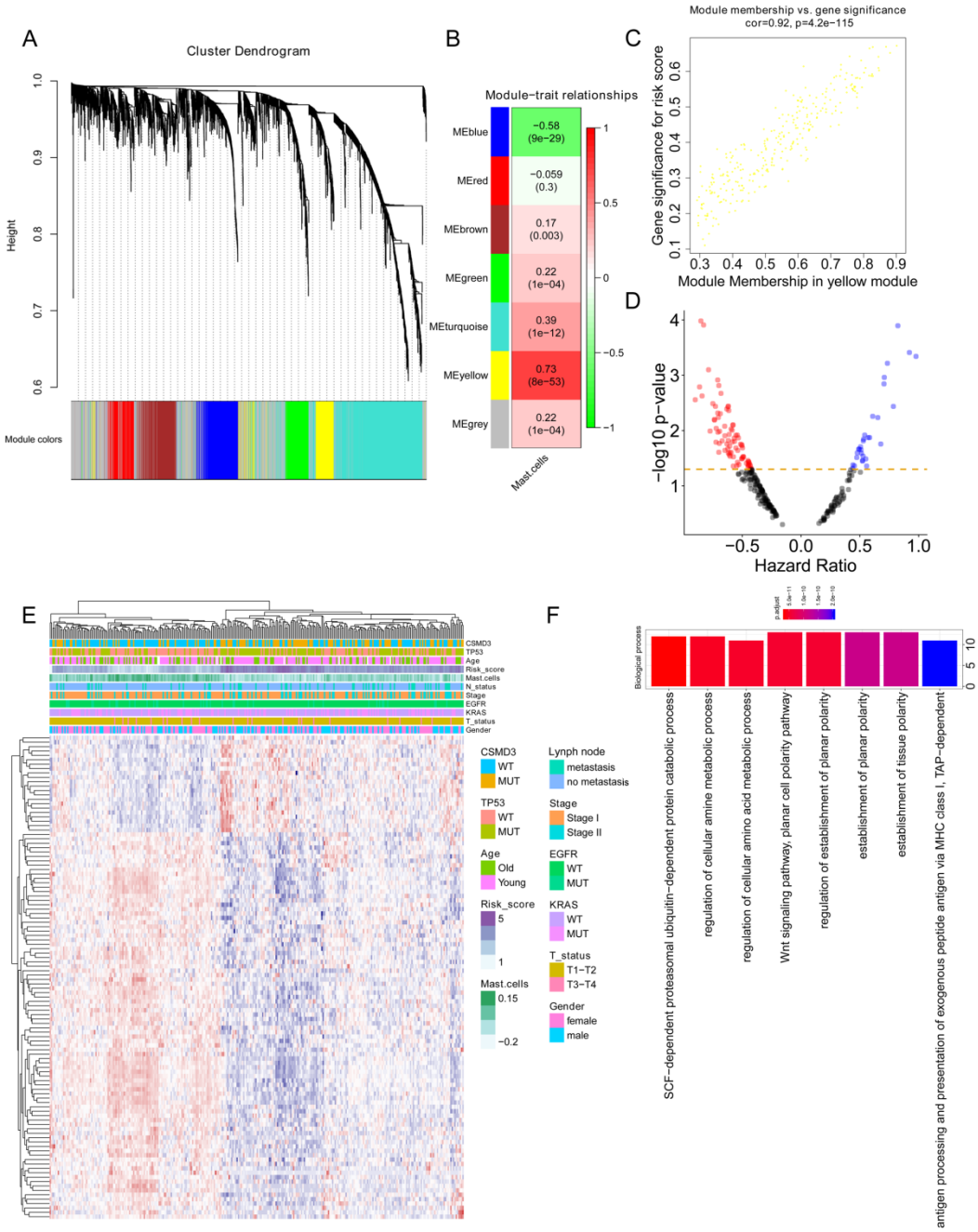
mechanism underlying the mast cells in early-stage LUAD. (A) PCA of the key mast cell-related genes. (B) The correlation between the prognostic signature-based risk score and the mast cell ssGSEA score in early-stage LUAD patients. (C) Univariate Cox analysis and Kaplan-Meier curves showed prolonged survival in patients with low-risk scores compared with patients with high-risk scores. (D) The correlation between the ssGSEA score of each hallmark gene and the risk score. (E) The correlation between the risk score and ssGSEA score in early-stage LUAD patients. (F) The risk score distribution in patients with wild-type or mutated TP53. P-value was calculated with Mann-Whitney U test. (G) A Sankey plot was used to reveal the correlation between mast cell scores, prognostic signature-based risk scores, immunotherapy response, and clinical outcome. (H&I) Patients who received adjuvant therapies, including chemo(radio)therapy and targeted therapy, with low-risk scores exhibited prolonged overall survival.

Fig. 6 Association of the immune signature with early-stage LUAD gene mutations. (A) The distribution of gene mutations correlated with the prognostic signature-based risk score. TP53 was the most important mutation according to the importance of ranking.

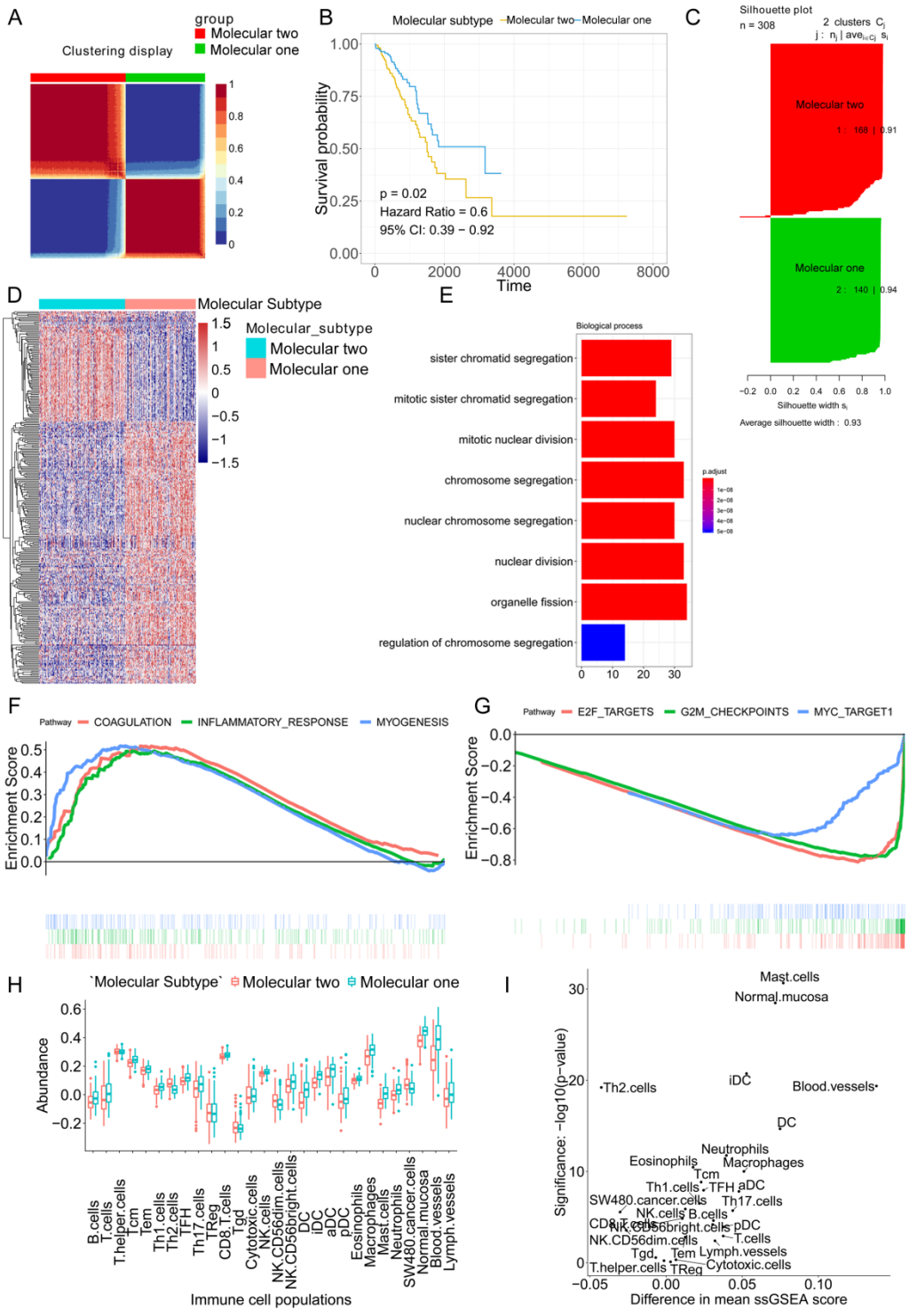
Fig. 7 Meta-analysis and external validation of the prognostic value of the mast cell-related signature. (A) Detailed information for the nine external validation cohorts. (B) A meta-analysis revealed the overall prognostic value of the mast cell-related signature.



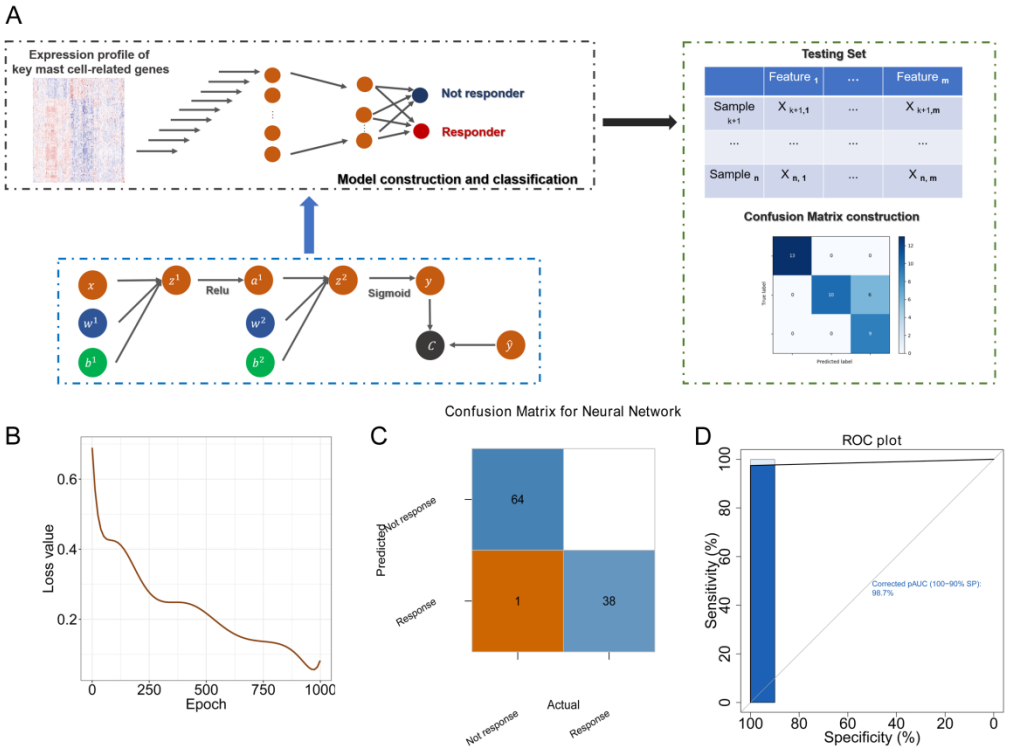
mol2_12670_f1.tif



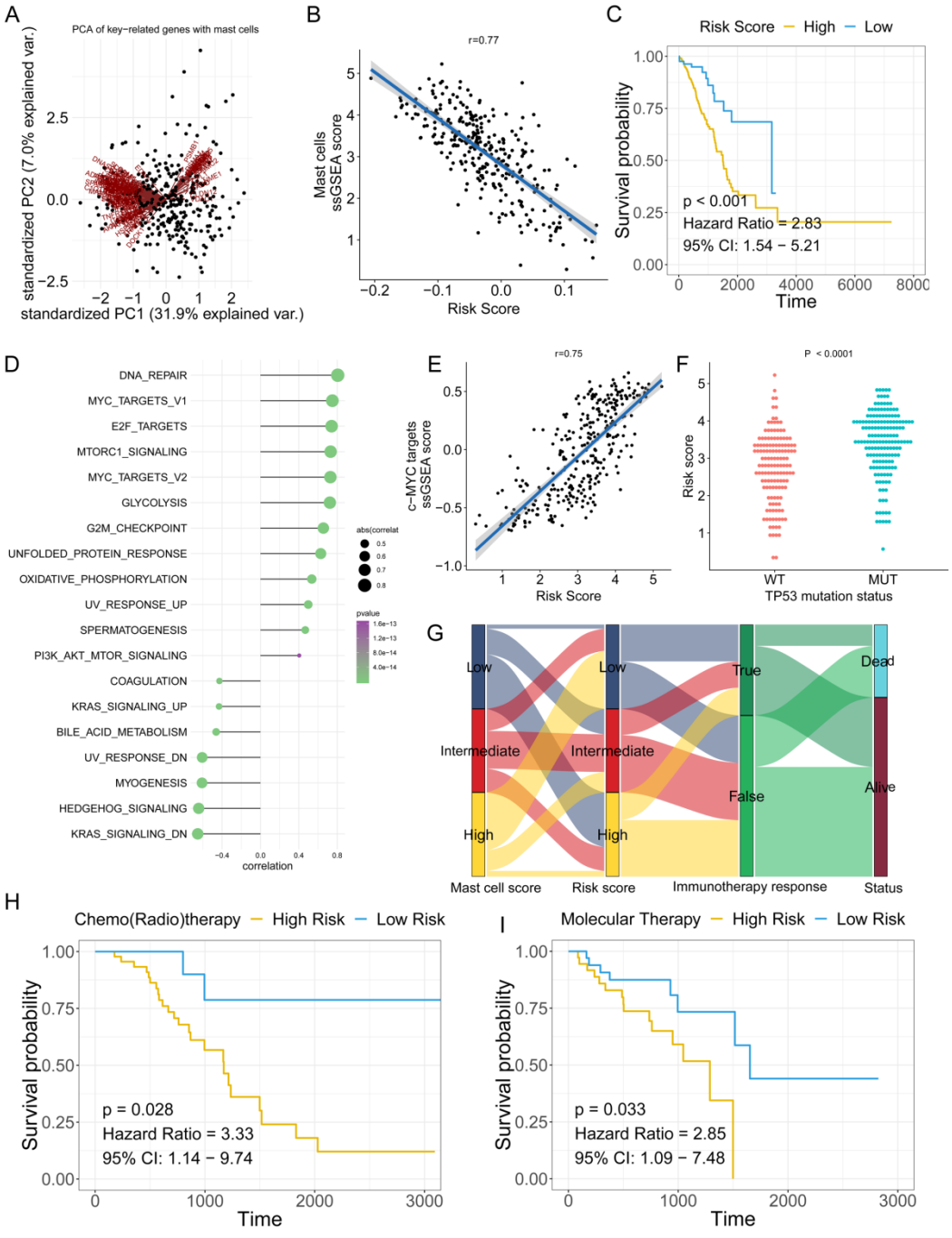
mol2_12670_f2.tif



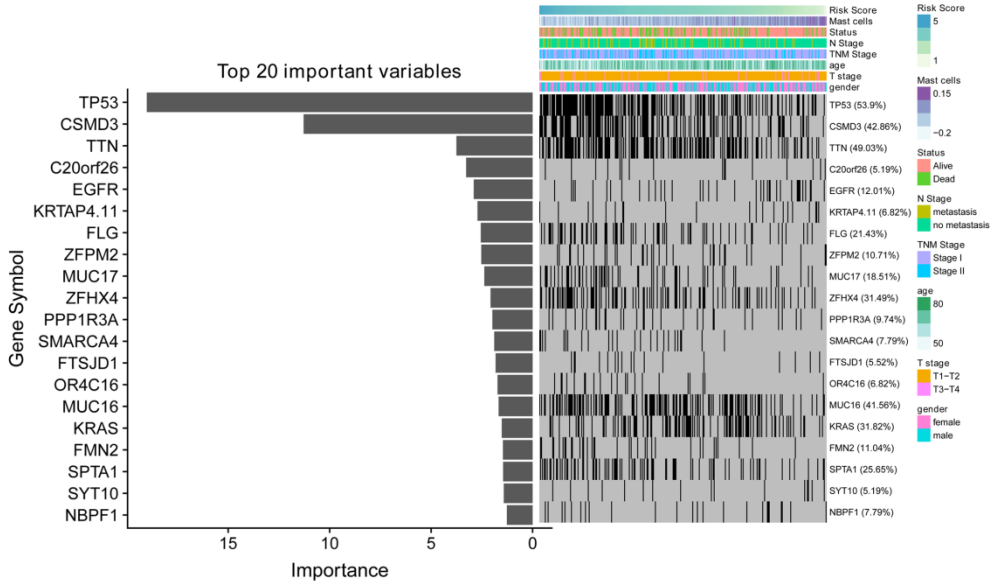
mol2_12670_f3.tif



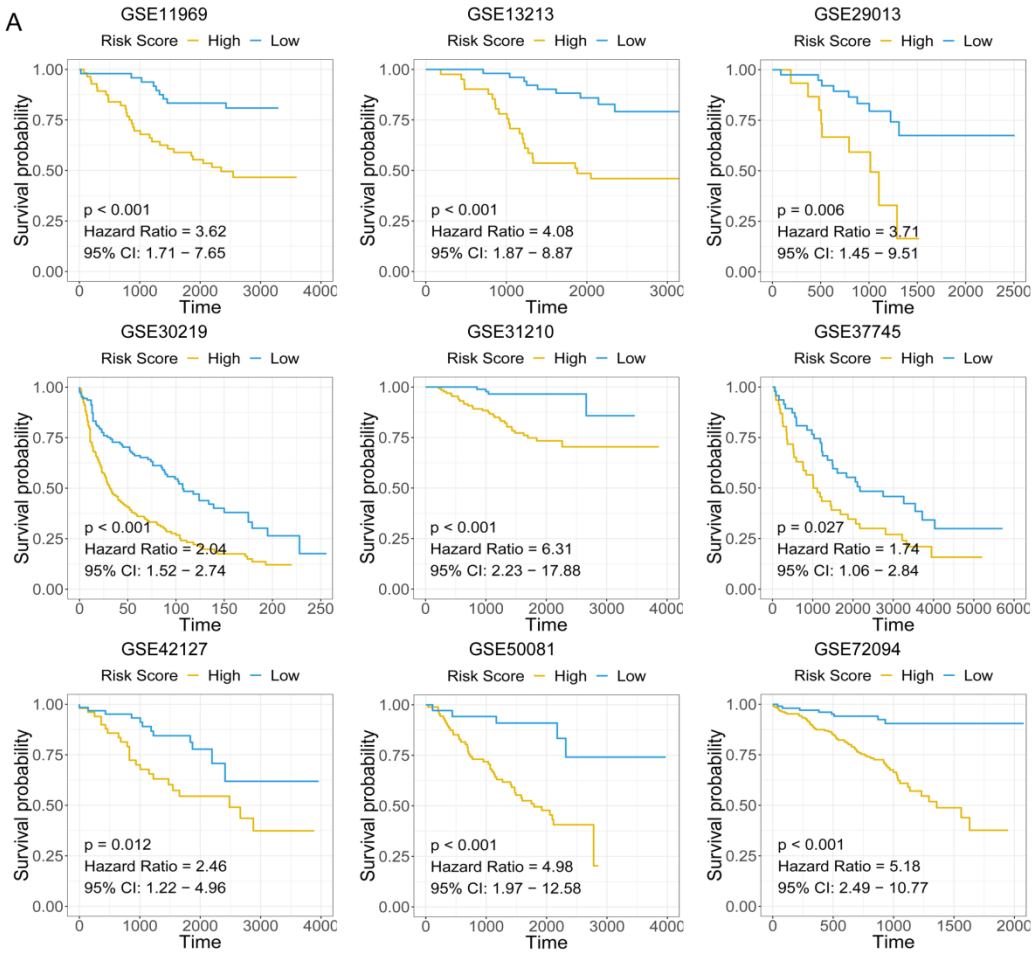
mol2_12670_f4.tif



mol2_12670_f5.tif



mol2_12670_f6.tif



mol2_12670_f7.tif

Table 1:

Patient information

Variable	Number
Gender (female/male)	166/142
TNM Stage (stage I/stage II)	212/96
Lymph node metastasis (positive/negative)	61/247
Age (>60/<=60/missing)	221/78/9
KRAS Status (WT/MUT)	210/98
EGFR Status (WT/MUT)	271/37
Smoking (no/yes/missing)	15/93/200