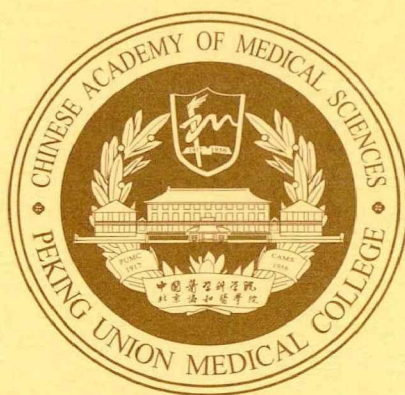


北京協和醫學院
中國醫藥科學院

博士研究生學位論文



北京協和醫學院研究生院

学校代码： 10023
学 号： B2021003075

博 士 学 位 论 文

（专业学位）

基于线粒体自噬和衰老表型相关基因构建肺腺癌预后模型并探讨 DSP 基因对肺腺癌细胞的作用研究

The construction of a prognostic model for lung adenocarcinoma using genes related to mitophagy and aging, and the study on the effect of the DSP gene on lung adenocarcinoma cells

所 院 ： 肿瘤医院

姓 名 ： 孟祥志

指导教师： 高禹舜

导师小组： 张良泽、邵康、冯飞跃

学科专业： 肿瘤学（胸外科专业型）

研究方向： 胸部肿瘤学

完成日期： 2024 年 5 月

目 录

前言.....	13
实验材料.....	15
一, 主要实验试剂.....	15
二, 主要实验器材.....	16
实验方法.....	19
一, 数据下载.....	19
二, LUAD 相关差异表达基因	22
三, 构建 MiAg DEGs 预后模型	22
四, 功能富集分析.....	23
五, 基因集富集分析.....	23
六, 高低风险分组间的基因集变异分析	23
七, 基于 TCGA 数据集的线粒体自噬和衰老表型评分计算	24
八, 免疫浸润分析.....	24
九, 预后临床相关性分析.....	24
十, 入组 LUAD 患者选择	25
十一, Real-Time qPCR 实验.....	26
十二, LUAD 中单细胞测序分析	28
十三, 多重荧光免疫组织化学	29
十四, 细胞培养及转染.....	29
十五, 蛋白免疫印记(Western Blot)检测.....	30
十六, CCK8 细胞增殖实验	31
十七, 平板克隆形成实验.....	31
十八, Transwell 实验.....	31
十九, 划痕实验.....	32
二十, 流式细胞学技术.....	33
二十一, 统计分析.....	33
实验结果.....	34
一, LUAD 数据集的标准化处理	34
二, 线粒体自噬和衰老相关基因在 LUAD 患者中的突变分析	36
三, LUAD 相关差异基因分析	39
四, MiAg DEGs 的相关性分析	41
五, MiAg DEGs 的表达差异分析	43
六, MiAg DEGs 功能富集分析	47
七, LUAD 数据集的基因集富集分析	50

八, MiAg DEGs 预后模型的构建以及高低风险分组间的差异分析	54
九, 高低风险组间差异表达基因的基因集变异分析	57
十, 线粒体自噬和衰老表型评分的构建与免疫治疗评估	59
十一, MiAg Scores 高低评分分组之间的免疫特征差异分析	61
十二, TCGA-LUAD 数据集多因素 Cox 预后模型的构建	63
十三, 本中心 LUAD 患者中构建预后模型的 MiAg DEGs 表达相关情况	66
十四, MiAg DEGs 在 LUAD 各类细胞中的表达和分布情况	67
十五, MiAg DEGs 相关蛋白在 LUAD 中的定位表达情况及临床相关性	71
十六, DSP 在 LUAD 细胞中表达下调	74
十七, DSP 在 LUAD 细胞中过表达和敲减效果验证	75
十八, DSP 抑制 LUAD 细胞的增殖	76
十九, DSP 促进 LUAD 细胞的凋亡	77
二十, DSP 抑制 LUAD 细胞的迁移和侵袭	78
讨论	81
小结	85
参考文献	86
肺腺癌预后模型的现状和研究进展(文献综述)	90
参考文献	95
基金资助	98

表索引

表 1. TCGA 数据集中 LUAD 患者的临床特征 19

表 2. LUAD 数据集列表 21

表 3. 线粒体自噬和衰老相关差异基因的功能富集分析 48

表 4. TCGA-LUAD 数据集中的基因集富集分析 53

表 5. MiAg DEGs 预后模型中高低评分分组间的基因组变异分析 58

表 6. TCGA-LUAD 数据集中 COX 回归分析 65

图索引

图 1. 线粒体自噬和衰老相关基因构建 LUAD 预后模型的技术路线图	25
图 2. TCGA-LUAD 数据集的标准化处理	35
图 3. GEO 中 LUAD 数据集的标准化处理	36
图 4. 线粒体自噬和衰老相关基因在肺腺癌患者中的突变分析	38
图 5. LUAD 中差异基因分析与线粒体自噬和衰老相关差异基因热图	40
图 6. TCGA-LUAD 和 GSE10072 数据集中 MiAg DEGs 的相关性分析	42
图 7. GSE30219 数据集中 MiAg DEGs 的相关性分析	43
图 8. TCGA-LUAD 数据集中 MiAg DEGs 的表达差异分析	45
图 9. GSE10072 数据集中 MiAg DEGs 的表达差异分析	46
图 10. GSE30219 数据集中 MiAg DEGs 的表达差异分析	47
图 11. MiAg DEGs 功能富集分析	50
图 12. LUAD 数据集的基因集富集分析	51
图 13. MiAg DEGs 预后模型的构建与高低风险评分分组间的差异分析	56
图 14. 高低风险组间 DEGs 的基因集变异分析	58
图 15. MiAg Scores 的构建与高低风险评分分组间的差异分析	61
图 16. TCGA-LUAD 数据集 MiAg Scores 高低评分分组的免疫浸润分析	63
图 17. 高低评分组 MiAg Scores 的临床预后模型构建	66
图 18. 本中心 LUAD 患者中 5 个 MiAg DEGs 的表达情况	67
图 19. EMTAB6149 数据集中 MiAg DEGs 表达和分布的情况	69
图 20. GSE127465 数据集中 MiAg DEGs 表达和分布的情况	70
图 21. GSE143423 数据集中 MiAg DEGs 表达和分布的情况	71
图 22. MiAg DEGs 相关蛋白在 LUAD 中的定位表达情况	73
图 23. LUAD 中不同年龄和性别因素中 MiAg DEGs 的表达变化	74
图 24. LUAD 中不同临床特征中 MiAg DEGs 表达变化	74
图 25. DSP 在 LUAD 细胞系中表达降低	75
图 26. DSP 在 LUAD 细胞中过表达和敲减验证	76
图 27. DSP 对 LUAD 细胞增殖能力的影响	77

图 28. DSP 对 LUAD 细胞凋亡的影响.....	78
图 29. 划痕实验检测 DSP 对 H1975 细胞迁移的影响。.....	79
图 30. 划痕实验检测 DSP 对 A549 细胞迁移的影响。.....	79
图 31. 划痕实验检测 DSP 对 H1395 细胞迁移的影响.....	79
图 32. DSP 对 LUAD 细胞侵袭能力的影响.....	80

中英文缩略词

英文缩写	英文全称	中文名称
LUAD	Lung adenocarcinoma	肺腺癌
TCGA	The cancer genome atlas	肿瘤基因组计划
GEO	Gene Expression Omnibus	基因表达综合数据库
MSigDB	Molecular Signatures Database	分子签名数据库
MiAg DEGs	Mitophagy and Aging related differentially expressed genes	线粒体自噬和衰老相关差异表达基因
Lasso	Least absolute shrinkage and selection operator	最小绝对收缩与选择算子
GO	Gene Ontology	基因本体论
BP	biological process	生物过程
MF	molecular function	分子功能
CC	cellular component	细胞成分
GSEA	Gene Set Enrichment Analysis	基因集富集分析
GSVA	Gene Set Variation Analysis	基因集变异分析
ssGSEA	single-sample gene-set enrichment analysis	单样本基因集富集分析
MiAg Scores	Mitophagy and Aging related scores	线粒体自噬和衰老表型相关评分
DCA	Decision curve analysis	决策曲线分析
Real Time-qPCR	Real Time Quantitative Polymerase Chain Reaction	实时荧光定量聚合酶链式反应
OS	Overall Survival	总生存

摘要

目的：线粒体自噬和衰老是肿瘤发生发展过程中的重要病理生理机制。然而线粒体自噬和衰老相关基因在肺腺癌中的预后价值以及其对肺腺癌发生发展的影响目前尚不清楚。我们拟通过肺腺癌中线粒体自噬和衰老双表型相关的差异表达基因构建肺腺癌预后模型并探讨其中的 DSP 基因对肺腺癌细胞增殖、凋亡、侵袭和迁移的影响。

方法：首先，我们使用"Mitophagy"和"Aging"作为关键词，在 GeneCards 数据库、PubMed 网站以及分子签名数据库中筛选与线粒体自噬和衰老表型相关的基因。接着从 TCGA 数据库中下载了肺腺癌数据集作为测试集，从 GEO 数据库下载 GSE10072、GSE30219、GSE3141 三个数据集作为验证集。通过基因表达差异分析，识别出与线粒体自噬和衰老相关的差异表达基因。

其次，我们在 TCGA 肺腺癌数据集中使用 Lasso 回归方法构建了预后模型。同时对线粒体自噬和衰老相关的差异表达基因进行了 GO 富集分析、基因集富集分析和基因集变异分析。此外，还利用单样本基因集富集分析算法和 CIBERSORT 算法来评估每个样本的线粒体自噬和衰老相关评分以及和免疫细胞的相关性。进一步结合这些评分和临床变量进行了单因素和多因素 Cox 回归分析，并构建基于多因素 Cox 回归的临床预测模型。通过对肺腺癌单细胞测序数据的分析评估线粒体自噬和衰老相关差异基因在肺腺癌中不同细胞类型上的表达和定位情况。利用实时荧光定量聚合酶链式反应技术和多重荧光免疫组织化学技术在肺腺癌患者样本中检测这些基因的 mRNA 表达水平和蛋白表达情况，并分析了它们与患者的临床病理特征和预后的相关性。

最后，通过 CCK-8、平板克隆、流式细胞学、划痕和细胞侵袭实验等方法进一步分析肺腺癌细胞来源的预后相关基因 DSP 对肺腺癌细胞系 H1975、A549 及 H1395 增殖、凋亡、侵袭和迁移的影响。

结果：

1，共筛选出了 560 个与线粒体自噬和衰老相关的基因。利用 TCGA-LUAD、GSE10072 和 GSE30219 数据集进行差异表达基因分析，最终确定了 7 个与肺腺癌中线粒体自噬和衰老相关的差异表达基因：CAV1、DSG2、DSP、MYH11、NME1、PAICS、PLOD2。

2，GO 富集分析显示，这些基因主要富集于心室肌细胞动作电位调节、细胞-细胞接触区和细胞-细胞粘附介质活性等生物学过程。

3, 基因集富集分析表明, 它们在 MAPK 通路和 NF κ B 通路的激活中显著富集。

4, 通过 Lasso 回归分析, 确定了 CAV1、DSG2、DSP、MYH11、PLOD2 这 5 个基因用于构建肺腺癌预后模型, 该模型在预测肺腺癌患者生存结局方面显示出较高的准确性。

5, 线粒体自噬和衰老基因相关评分在不同样本预后结局间具有显著差异。8 种免疫细胞的浸润丰度在高低评分组间具有显著差异, 且多数免疫细胞的浸润丰度呈负相关性。

6, 多因素 Cox 预后模型表明, 线粒体自噬和衰老相关基因评分与肺腺癌患者预后显著相关, 且预测效果随时间增长而提高。

7, 实时荧光定量聚合酶链式反应结果显示, 和癌旁组织相比, DSG2 在肺腺癌组织中的表达显著上调。

8, 单细胞测序数据集分析揭示, DSG2 和 DSP 主要在肺腺癌恶性肿瘤细胞中表达。

9, 多重荧光免疫组织化学染色结果显示 DSG2 和 DSP 主要在肺腺癌细胞中表达。在临床相关性分析中, DSG2 和 DSP 在不同肿瘤分期和分化水平上的表达具有统计学差异, 特别是 DSP 在低分化肺腺癌、死亡患者组中显著降低, 而 DSG2 与之相反。

10, 在肺腺癌细胞系 (H1975、A549 及 H1395) 中进行 DSP 的过表达和敲减, 通过体外细胞实验证实, 过表达 DSP 抑制了肺腺癌细胞的增殖、侵袭和迁移, 并促进了肺腺癌细胞的凋亡, 而敲减 DSP 逆转了以上结果。

结论: 本研究通过分析首次确定了肺腺癌中线粒体自噬和衰老相关的 5 个关键差异表达基因。这些基因在肺腺癌发生发展中的表达差异和生物学功能揭示了其潜在的临床价值, 特别是在预后模型构建和肺腺癌细胞功能方面, 为肺腺癌的治疗提供了新的见解和潜在的治疗靶点。

关键词: 线粒体自噬, 衰老, 肺腺癌, 预后模型, DSP

Abstract

Objective: Mitophagy and aging are important pathophysiological mechanisms in the development and progression of tumors. However, the prognostic value of genes related to mitophagy and aging in lung adenocarcinoma, as well as their impact on the development and progression of lung adenocarcinoma, remains unclear. We aim to construct a prognostic model for lung adenocarcinoma based on differentially expressed genes related to both mitophagy and aging phenotypes in lung adenocarcinoma and to explore the effects of the DSP gene on the proliferation, apoptosis, invasion, and migration of lung adenocarcinoma cells.

Method: Firstly, we used "Mitophagy" and "Aging" as keywords to filter genes related to mitophagy and aging phenotypes from the GeneCards database, PubMed website, and the Molecular Signatures Database. Subsequently, lung adenocarcinoma datasets were downloaded from the TCGA database as the test set, and datasets GSE10072, GSE30219, and GSE3141 were downloaded from the GEO database as the validation set. Through differential gene expression analysis, genes differentially expressed in relation to mitophagy and aging were identified.

Secondly, we constructed a prognostic model in the TCGA lung adenocarcinoma dataset using the Lasso regression method. We also performed GO enrichment analysis, gene set enrichment analysis, and gene set variation analysis on the differentially expressed genes related to mitophagy and aging. In addition, the single-sample gene set enrichment analysis algorithm and the CIBERSORT algorithm were used to evaluate the scores related to mitophagy and aging in each sample, as well as their correlation with immune cells. Further, these scores and clinical variables were combined in univariate and multivariate Cox regression analyses, and a clinical prediction model based on multivariate Cox regression was constructed. The expression and localization of mitophagy and aging-related differentially expressed genes in different cell types of lung adenocarcinoma were assessed by analyzing single-cell sequencing data of lung adenocarcinoma. The mRNA and protein expression levels of these genes in lung adenocarcinoma were detected using real-time quantitative PCR technology and multiplex immunohistochemistry, and their correlation with the patients' clinicopathological characteristics and prognosis was analyzed.

Lastly, the effect of the prognostic-related gene DSP derived from lung

adenocarcinoma cells, on the proliferation, apoptosis, invasion, and migration of lung adenocarcinoma cell lines H1975, A549, and H1395 was further analyzed using methods such as CCK-8, colony formation assay, flow cytometry, cell migration, and cell invasion assays.

Results:

1, A total of 560 genes related to mitophagy and aging were identified. Through differential gene expression analysis using the TCGA-LUAD, GSE10072, and GSE30219 datasets, seven differentially expressed genes related to mitophagy and aging in lung adenocarcinoma were ultimately identified: CAV1, DSG2, DSP, MYH11, NME1, PAICS, PLOD2.

2, GO enrichment analysis showed that these genes were mainly enriched in biological processes such as regulation of ventricular cardiac muscle cell action potential, cell-cell junctions, and cell-cell adhesion mediator activity.

3, Gene set enrichment analysis indicated that they were significantly enriched in the activation of the MAPK pathway and NF- κ B pathway.

4, Through LASSO regression analysis, CAV1, DSG2, DSP, MYH11, and PLOD2 were determined to be used for constructing a prognostic model for lung adenocarcinoma, which showed high accuracy in predicting the survival outcomes of patients with lung adenocarcinoma.

5, There was a significant difference in mitophagy and aging gene-related scores between different sample prognostic outcomes. The infiltration abundance of eight types of immune cells had significant differences between high and low score groups, with most showing a negative correlation in infiltration abundance.

6, The multivariate Cox prognostic model indicated that the scores related to mitophagy and aging genes were significantly correlated with the prognosis of patients with lung adenocarcinoma, and the prediction effect improved over time.

7, Real-time fluorescent quantitative PCR results showed that DSG2 was significantly upregulated in lung adenocarcinoma tissues compared to adjacent non-cancerous tissues.

8, Single-cell sequencing dataset analysis revealed that DSG2 and DSP were mainly expressed in malignant tumor cells of lung adenocarcinoma.

9, Multiplex immunohistochemistry staining results showed that DSG2 and DSP were primarily expressed in lung adenocarcinoma cells. In the analysis of clinical relevance, the expression of DSG2 and DSP showed statistical differences across different tumor stages

and differentiation levels, especially DSP, which was significantly reduced in poorly differentiated lung adenocarcinoma and in the deceased patient group, while DSG2 showed the opposite trend.

10, By constructing DSP overexpressing and silenced lung adenocarcinoma cell lines (H1975, A549, and H1395) using plasmids and siRNA, subsequent in vitro cell experiments indicated that overexpressing DSP inhibited the proliferation, invasion, and migration of lung adenocarcinoma cells and promoted apoptosis, while silencing DSP reversed these results.

Conclusion: This study has identified five key differentially expressed genes related to mitophagy and aging in lung adenocarcinoma for the first time through analysis. The differential expression and biological functions of these genes in the development of lung adenocarcinoma reveal their potential clinical value, especially in the construction of prognostic models and the functional aspects of lung adenocarcinoma cells, providing new insights and potential therapeutic targets for the treatment of lung adenocarcinoma.

Keywords: Mitophagy, Aging, Lung Adenocarcinoma, Prognostic Model, DSP

前言

肺癌是一种具有高发病率和高死亡率的肿瘤[Bray et al, 2018]。在肺癌的各种病理亚型中,非小细胞肺癌的比例接近 80-85%,其主要有两种亚型,即肺腺癌(Lung adenocarcinoma, LUAD)和肺鳞癌,其中 LUAD 是最常见的类型[Ettinger et al, 2015]。尽管近年来诊疗技术不断进步,但 LUAD 患者的预后仍然很差,尤其是癌症晚期患者,五年生存率仅为 10-15%[Siegel et al, 2012]。虽然手术、放疗和化疗是治疗 LUAD 最常用的方法,但它们对延长 LUAD 患者生存期的作用有限[Zhou et al, 2018]。在临床工作中,相当一部分 LUAD 患者在初次就诊时通常就被诊断为晚期[Zeng et al, 2020],预后很差。近年来,随着基因测序等分子生物学技术的普及和新表型的不断被发现,使得基于此建立一种可以预测 LUAD 患者预后的模型及开发新的预后生物标志物作为靶点来治疗 LUAD 成为可能。

线粒体自噬是自噬的一种特殊形式,主要负责清除受损或多余的线粒体,以维持线粒体数量和质量的稳定,从而维持细胞的能量供应和平衡[Wang, Y. et al, 2020]。近年来,研究显示线粒体自噬参与肿瘤的发生发展[Sharma et al, 2021]。其中,肺癌的发生也与一系列线粒体自噬相关基因有关[Denisenko et al, 2021]。例如:低 ATG5 表达会降低 RAS 突变肺癌细胞系的肿瘤生长[Guo et al, 2011],低 PINK1 会改变肿瘤的能量代谢过程、抑制非小细胞肺癌对糖酵解相关药物的敏感性[Dai et al, 2019]。

除此之外,近年来的研究表明有许多基因和信号通路在机体衰老的过程中发生了改变。目前,大规模的人群转录组测序分析已经明确了部分与衰老相关的基因[Peters et al, 2015],且研究已经证实这些基因参与了癌症的发生发展过程。如在三阴性乳腺癌和前列腺癌中,衰老相关基因与较弱的免疫浸润和较差的患者生存率有关[Chen et al, 2020],这表明衰老相关基因可能与肿瘤的免疫治疗反应相关。研究人员在黑色素瘤和非小细胞肺癌患者的基因组突变数据中观察到,在这两种肿瘤中存在衰老突变基因的患者预后较差[Chong et al, 2021]。

随着年龄的增长,细胞的线粒体自噬能力会逐渐下降,导致受损线粒体的积累,从而加速细胞老化和组织功能的衰退。研究表明,激活线粒体自噬可以延缓老化过程,减少与年龄相关的疾病风险,说明线粒体自噬在调节衰老和维持细胞稳态中扮演着重要角色,也说明二者之间存在着密切的联系[Wang, Y. et al, 2020]。

上述研究表明,线粒体自噬和衰老相关基因之间存在密切联系且在肿瘤的发生发展中发挥着重要作用。虽然有研究分别利用线粒体自噬和衰老相关基因构建了预测模型,分别用于预测 LUAD 患者的预后和免疫浸润[Dai et al, 2022; Zhang et al, 2022],也取得了较好的预测效果,但这些研究所涉及的基因数量偏多,不利于确定具体的作用靶点。并且同时利用线粒体自噬和衰老双表型相关基因构建 LUAD 预后

模型的相关研究目前尚未见报道。因此,在本论文中,我们利用生物信息学技术对公共数据库进行了数据分析,确定了与线粒体自噬和衰老双表型相关的差异基因并通过这些基因结合临床病理因素构建了 LUAD 的临床预后模型,并分析了线粒体自噬和衰老双表型相关的基因与 LUAD 预后和免疫浸润等的潜在相关性。最后我们在本院自己患者的数据上对线粒体自噬和衰老双表型相关的基因进行了表达和定位的验证,发现 DSP 基因表达和定位于肺腺癌细胞上并与肺腺癌的分化及临床病理分期密切相关,因此后续实验中就 DSP 基因对 LUAD 细胞的作用进行了初步探索。

综上,我们的研究可能有助于阐明线粒体自噬和衰老双表型相关差异基因在 LUAD 发病和患者预后过程中的作用,并为开发 LUAD 新的预后标志物和治疗靶点提供新的方向和理论依据。

实验材料

一，主要实验试剂

试剂	厂家	货号
无水乙醇	国药集团化学试剂有限公司	100092183
环保型脱蜡液	Wuhan Servicebio Technology Co., Ltd.	G1128
柠檬酸抗原修复液 (PH6.0)	Wuhan Servicebio Technology Co., Ltd.	G1202
EDTA 抗原修复液 (PH9.0)	Wuhan Servicebio Technology Co., Ltd.	G1203
EDTA 抗原修复液 (PH8.0)	Wuhan Servicebio Technology Co., Ltd.	G1206
PBS 缓冲液	Wuhan Servicebio Technology Co., Ltd.	G0002
组织自发荧光淬灭剂	Wuhan Servicebio Technology Co., Ltd.	G1221
牛血清白蛋白 BSA	Wuhan Servicebio Technology Co., Ltd.	GC305010
兔血清	Wuhan Servicebio Technology Co., Ltd.	G1209
3%过氧化氢消毒液	山东安捷高科消毒科技有限公司	
DAPI 染色试剂	Wuhan Servicebio Technology Co., Ltd.	G1012
抗荧光淬灭封片剂	Wuhan Servicebio Technology Co., Ltd.	G1401
无菌 PBS	Wuhan Servicebio Technology Co., Ltd.	G4250
TBST	Wuhan Servicebio Technology Co., Ltd.	G0004
CAV-1 抗体	武汉三鹰生物技术有限公司	16447-1-AP
DSG2 抗体	武汉三鹰生物技术有限公司	21880-1-AP
DSP 抗体	Wuhan Servicebio Technology Co., Ltd.	GB113491
MYH11 抗体	Wuhan Servicebio Technology Co., Ltd.	GB15011
PLOD2 抗体	武汉三鹰生物技术有限公司	21214-1-AP
山羊抗兔 IgG 抗体	BioVision 公司	6927-100
蛋白提取液	Medical Discovery Leader	MDL91201
蛋白酶抑制剂	Medical Discovery Leader	MD912893
BCA 蛋白浓度测定 kit	Medical Discovery Leader	MD913053
SDS-PAGE 预制胶套装 kit	Medical Discovery Leader	MD911919
蛋白分子量 marker	苏州博奥龙科技有限公司	BP06193S

二抗	Medical Discovery Leader	MD912565
蛋白裂解液	碧云天生物技术有限公司	P0013
Actin 抗体	Affinity	
DSG2 抗体	Proteintech	21880-1-AP
SYBR Green Mix(2X)	UE	S2008
cDNA 合成试剂盒	Takara	RR047A
Trizol	Genebetter	R011-100
引物	南京金斯瑞生物	
Cell Counting Kit-8	Fluorescence	DCM7122
结晶紫染液	Medical Discovery Leader	MD911626
PBS	Medical Discovery Leader	MD911702
甲醇	西陇科学	1710091
胰酶	Gibco	25200-072
细胞凋亡检测试剂盒	四正柏生物	FXP018

二，主要实验器材

名称	厂家	型号
脱水机	意大利迪佩斯 (DIPATH)	Donatello
包埋机	武汉俊杰电子有限公司	JB-P5
抗原修复盒	Wuhan Servicebio Technology Co., Ltd.	WGXFHA0008
病理切片机	上海徕卡仪器有限公司	RM2016
微量加样器	大龙公司, 中国	微量加样器
冻台	武汉俊杰电子有限公司	JB-L5
钟摆摇床	Wuhan Servicebio Technology Co., Ltd.	SYC-Z100
组织摊片机	浙江金华市科迪仪器设备有限公司	KD-P
电热恒温水浴槽	上海一恒 恒温设备厂	HWS-24
烤箱	上海慧泰仪器制造有限公司	DHG-9140A
涡旋混合器	Wuhan Servicebio Technology Co., Ltd.	MV-100

冰箱	青岛海尔股份有限公司	BCD-192TGN
NC 膜 (0.22μm)	GVS, 美国	1215458
载玻片	Wuhan Servicebio Technology Co., Ltd.	G6012
掌上离心机	Wuhan Servicebio Technology Co., Ltd.	MS6000
盖玻片	江苏世泰实验器材有限公司	10212432C
3MM 滤纸	Whatman, 英国	3030861
微波炉	格兰仕微波炉电器有限公司	P70D20TL-P4
避光湿盒	Wuhan Servicebio Technology Co., Ltd.	SIB-20F
移液器	Wuhan Servicebio Technology Co., Ltd.	SPIP
组化笔	Wuhan Servicebio Technology Co., Ltd.	G6100
正置荧光显微镜	日本尼康	NIKON ECLIPSE C1
脱色摇床	海门其林贝尔仪器制造公司	TS-100
扫描仪	3DHISTECH	Pannoramic MIDI
微量移液器	大龙公司, 中国	
实时荧光定量 PCR 仪	Longene	Q2000B
电泳仪	北京百晶生物技术有限公司	BG-subMIDI
荧光定量 PCR 管	Axygen	SQ9604
超微量分光光度计	北京美林恒通	SMA1000
1ml 无 RNase 袋装吸头	Axygen	
200ul 无 RNase 袋装吸头	Axygen	
10ul 无 RNase 袋装吸头	Axygen	
低温离心机	LABGIC, 中国	CF1524R
SDS-PAGE 电泳系统	BIO-Rad, 美国	mini-PROTEAN® Tetra Cell with mini Trans-Blot® Module
凝胶成像系统	UVP, 美国	GelDoc-It310
化学发光成像系统	CLINX 勤翔, 中国	ChemiScope6100
离心机	恒诺仪器	AXTGL16M

酶标仪	BIO-RAD	550
移液器	Eppendorf	—
显微镜	NOVEL	XS-2100
Transwell 小室	COSTar	3422
流式细胞仪	BENM DIAG	BeamCyte-1026

实验方法

一、数据下载

我们通过 R 语言中的 TCGAbiolinks 包[Colaprico et al, 2016]从肿瘤基因组计划（The cancer genome atlas, TCGA; <https://portal.gdc.cancer.gov/>）中下载 LUAD 数据集（TCGA-LUAD）并作为测试集进行分析，排除缺少临床信息的数据样本后，共得到 598 例具有临床信息的 LUAD 样本的测序数据，其中包括 59 例癌旁正常样本和 539 例肺腺癌肿瘤样本的测序数据。TCGA-LUAD 中的数据被转换成了 FPKM 格式，关联的临床信息来源于 UCSC Xena 数据库 [Goldman et al, 2020] (<http://genome.ucsc.edu>)，详细的数据集描述参见表 1。

表 1. TCGA 数据集中 LUAD 患者的临床特征

临床特征	总计(%)
T 分期	
T1	176 (32.8%)
T2	292 (54.5%)
T3	49 (9.1%)
T4	19 (3.5%)
N 分期	
N0	350 (66.9%)
N1	97 (18.5%)
N2	76 (14.5%)
M 分期	
M0	365 (93.6%)
M1	25 (6.4%)
肿瘤分期	
Stage I	296 (55.7%)

临床特征	总计(%)
Stage II	125 (23.5%)
Stage III	84 (15.8%)
Stage IV	26 (4.9%)
性别	
女	289 (53.6%)
男	250 (46.4%)
年龄	
<= 65	257 (49.4%)
> 65	263 (50.6%)
总体生存期	
生存	347 (64.4%)
死亡	192 (35.6%)
疾病特异性生存期	
生存	383 (76.1%)
死亡	120 (23.9%)
无进展间隔期	
无进展	313 (58.1%)

另外我们还从基因表达综合数据库 (Gene Expression Omnibus, GEO) [Barrett et al, 2007]使用 R 包 GEOquery[Davis et al, 2007]下载了 LUAD 患者的表达谱数据集 GSE10072[Landi et al, 2008], GSE30219[Landi et al, 2008]和 GSE3141[Bild et al, 2006], 数据集 GSE10072, GSE30219 和 GSE3141 均来自 Homo sapiens, GSE10072 数据集的样本为 LUAD 患者的肿瘤组织和对应匹配的非肿瘤组织的芯片表达谱样本, 共 107 例样本: 包含 58 例 LUAD 肿瘤组织样本和 49 例对应匹配的非肿瘤组织样本,

数据平台为 GPL96 [HG-U133A]; GSE30219 数据集的样本为肺癌患者的肿瘤组织和对应匹配的非肿瘤肺组织芯片表达谱样本，共 307 例样本：包含 85 例 LUAD 肿瘤组织样本，39 例基底肺癌肿瘤组织样本，24 例肺类癌肿瘤组织样本，3 例肺大细胞癌肿瘤组织样本，56 例肺大细胞神经内分泌癌肿瘤组织样本，21 例小细胞肺癌肿瘤组织样本，61 例肺鳞状细胞癌肿瘤组织样本，4 例组织学未知肺癌肿瘤组织样本和 14 例对应匹配的非肿瘤肺组织样本，数据平台为 GPL570 [HG-U133_Plus_2]; GSE3141 数据集的样本为 LUAD 患者的肺癌肿瘤组织的芯片表达谱样本，共 111 例样本：包含 58 例 LUAD 肿瘤组织样本和 53 例肺鳞状细胞癌肿瘤组织样本，数据平台为 GPL570 [HG-U133_Plus_2]。

数据集探针名注释均使用芯片 GPL 平台文件。我们将数据集 GSE10072 中 58 例 LUAD 肿瘤组织样本和 49 例非肿瘤组织样本的表达谱数据纳入后续分析。数据集 GSE30219 中 85 例 LUAD 肿瘤组织样本和 14 例非肿瘤肺组织的表达谱数据纳入后续分析。数据集 GSE3141 中 58 例 LUAD 肿瘤组织样本的表达谱数据纳入后续分析。GSE10072，GSE30219，GSE3141 数据集作为验证集进行后续验证，具体数据集信息见表 2。

表 2. LUAD 数据集列表

	TCGA-LUAD	GSE10072	GSE30219	GSE3141
数据平台	TCGA	GPL96	GPL570	GPL570
种属	人	人	人	人
组织	肺腺癌	肺腺癌	肺肿瘤	肺肿瘤
LUAD 组织	539 例	58 例	85 例	58 例
非肿瘤组织	59 例	49 例	14 例	/

GeneCards 数据库[Stelzer et al, 2016] (<https://www.genecards.org/>) 提供了有关人类基因的全面信息，我们以 “Mitophagy”一词作为搜索关键词，仅保留“Protein Coding”且“Relevance score > 1.000”的线粒体自噬相关基因。此外，以“Mitophagy”作为搜索关键词在 PubMed 网站上已发表文献[Chen et al, 2021; Wang et al, 2021; Zhuo et al, 2021]中获取线粒体自噬相关基因集。其中，GeneCards 数据库共包含 759 个线粒体自噬相关基因，PubMed 网站已发表的文献中共包含 43 个线粒体自噬相关基因。此外，以“Mitophagy”作为搜索关键词在分子签名数据库（Molecular Signatures Database, MSigDB; <https://www.gsea-msigdb.org/>) [Liberzon et al, 2015]得到 GObp positive regulation of mitophagy in response to mitochondrial depolarization, mitophagy, PINK1-PRKN mediated mitophagy 和 receptor mediated mitophagy 参考基因集中共 34

个线粒体自噬相关基因，经过合并去重复后共获得了 761 个线粒体自噬相关基因。

另外，我们以“Aging”一词作为搜索关键词，在 GeneCards 数据库中仅保留“Protein Coding”且“Relevance score > 1.000”，共得到衰老相关基因 9364 个。此外，以“Aging”作为搜索关键词在 PubMed 网站得到已发表文献[He et al, 2022]中的衰老相关基因 307 个。此外，以“Aging”作为搜索关键词在 MSigDB 中得到 GObp aging, GObp multicellular organism aging, GObp muscle atrophy, GObp skeletal muscle atrophy, GObp striated muscle atrophy, caloric restriction and aging, aging premature, NAD metabolism sirtuins and aging, somatroph axis GH and its relationship to dietary restriction and aging, GObp replicative senescence 和 GObp cellular senescence 参考基因集中共 327 个衰老相关基因，经过合并去重复后共获得了 9404 个衰老相关基因。

我们将 9404 个衰老相关基因和 761 个线粒体自噬相关基因合并取交集后，并且去除在数据集 TCGA-LUAD, GSE10072, GSE30219 探针替换过程中缺少的部分基因后，最终共获得了 560 个肺腺癌中线粒体自噬和衰老相关基因。

二，LUAD 相关差异表达基因

为识别出差异基因在 LUAD 中的潜在作用机制和相关生物学特征和分子通路，我们先使用 limma 包[Ritchie et al, 2015]对数据集 TCGA-LUAD，数据集 GSE10072 和数据集 GSE30219 进行标准化处理，通过差异分析，从 3 个 LUAD 数据集中识别出了癌症组与正常组之间的差异表达基因。采用 $|\log FC| > 1$ 和 $P_{adj} < 0.05$ 作为筛选标准，选出了用于进一步研究的差异表达基因。在这些基因中， $\log FC > 1$ 且 $P_{adj} < 0.05$ 的为表达上调基因，而 $\log FC < -1$ 且 $P_{adj} < 0.05$ 的为表达下调基因。

为得到 LUAD 的线粒体自噬和衰老相关差异表达基因 (Mitophagy and Aging related differentially expressed genes, MiAg DEGs)，我们首先将 TCGA-LUAD 数据集，GSE10072 数据集和 GSE30219 数据集差异分析得到的全部差异表达基因取交集并绘制韦恩图获得数据集的共同差异表达基因，然后将三个数据集的共同差异表达基因与 560 个线粒体自噬和衰老相关基因取交集，通过使用 R 的 ggplot2 包创建火山图和 pheatmap 包绘制热图并展示了差异分析的结果。此外，对交集的差异表达基因进行了韦恩图的绘制。

三，构建 MiAg DEGs 预后模型

为了在 TCGA-LUAD 数据集中构建 MiAg DEGs 预后模型，在 TCGA-LUAD 数据集中，我们对数据进行了 Lasso 回归 (Least absolute shrinkage and selection operator regression)，采用 10 倍交叉验证和显著性水平 $p = 0.05$ 进行调整，确保了 1000 个训练周期的稳健性，以有效避免过拟合。我们进一步将 Lasso 回归的结果通过可视

化展示,并通过风险因子图详细分析了各样本根据风险评分的分类、生存结果以及各分类中差异表达基因的表达情况。我们将 Lasso 回归筛选后的 MiAg DEGs 在 GSE10072 数据集和 GSE30219 数据集中进行验证。

为了确定 TCGA-LUAD 数据集 LUAD 患者样本中 MiAg DEGs 预后模型高、低风险分组间相关的差异表达基因,通过 limma 包[Ritchie et al, 2015],我们对 TCGA-LUAD 数据集中的 LUAD 基因表达谱数据进行了差异性分析,采用 $|\log FC| > 1$ 和 $P_{adj} < 0.05$ 的标准识别在高低风险分组中差异表达的基因。表达上调的基因符合 $\log FC > 1$ 且 $P_{adj} < 0.05$,而表达下调的基因满足 $\log FC < -1$ 且 $P_{adj} < 0.05$ 。通过 R 包 ggplot2,我们将这些差异性结果以火山图形式呈现。

四, 功能富集分析

我们进行了基因本体论(Gene Ontology, GO)[Yu et al, 2020]富集分析,运用R包 clusterProfiler[Yu et al, 2012]对差异表达基因进行了评估,重点关注生物过程(BP)、分子功能(MF)和细胞组分(CC)。筛选条件设定为 $P_{adj} < 0.1$ 和FDR值(q_{value}) < 0.05 ,采用Benjamini-Hochberg (BH)法进行P值校正。

五, 基因集富集分析

我们还进行了基因集富集分析 (Gene Set Enrichment Analysis, GSEA)[Subramanian et al, 2005],首先将所有基因根据 $\log FC$ 值排序并分为高低两组,然后对 TCGA-LUAD 数据集中的基因进行富集分析。GSEA 的参数设置为:种子为 2023,计算次数为 1000 次,每个基因集包含的基因数介于 10 到 500 之间。采用 Benjamini-Hochberg (BH) 法进行 P 值校正,并从 MSigDB 中选取了 c2.cp.all.v2022.1.Hs.symbols.gmt 基因集进行分析,显著富集的基因集符合 $P_{adj} < 0.05$ 和 FDR 值(q_{value}) < 0.25 的标准。

六, 高低风险分组间的基因集变异分析

在进行基因集变异分析 (Gene Set Variation Analysis, GSVA) 时,我们采用了 Wang 等[Wang et al, 2011]描述的非参数和无监督方法。这一分析技术主要通过转换不同样本之间的基因表达量矩阵为样本间的基因集表达矩阵,用以评估不同生物通路的富集程度。对于 TCGA-LUAD 数据集中的高低风险分组,我们从 MSigDB 选取了“h.all.v7.4.symbols.gmt”基因集来执行 GSVA,以探究不同风险分组间通路的富集差异,并计算 GSVA 结果中的富集通路在 TCGA-LUAD 数据集高低表达分组间样本的功能差异, $P_{adj} < 0.05$ 被认为具有统计学意义。

七、基于 TCGA 数据集的线粒体自噬和衰老表型评分计算

单样本基因集富集分析 (single-sample gene-set enrichment analysis; ssGSEA) 算法可以量化数据集样本中每个基因的相对丰度。为识别出 MiAg DEGs 在 LUAD 中的潜在作用机制和相关生物学特征及通路情况, 我们使用 R 语言 GSVA 包 [Hanzelmann et al, 2013] 通过 ssGSEA 算法依据 TCGA 数据集中每个样本的 MiAg DEGs 的表达矩阵计算了数据集中每个样本的线粒体自噬和衰老表型相关评分 (Mitophagy and Aging related scores, MiAg Scores)。然后我们将 TCGA-LUAD 数据集 LUAD 患者样本 MiAg Scores 根据中位数分成高、低评分两组。

八、免疫浸润分析

利用 CIBERSORT 算法 [Chen et al, 2018], 一种基于线性支持向量回归原理的技术, 我们对 TCGA-LUAD 数据集中的转录组表达矩阵进行了反卷积分析。此方法允许我们从混合细胞样本中估算出各种免疫细胞的比例和丰度。我们上传了具有不同 MiAg Scores 分组的矩阵数据至 CIBERSORT 网站 (<https://cibersortx.stanford.edu/>), 并使用 LM22 标记基因矩阵来筛选出免疫细胞富集得分高于零的样本, 从而获得并展示了详细的免疫细胞浸润丰度矩阵。TCGA-LUAD 数据集 MiAg Scores 高低分组样本间免疫细胞的浸润丰度差异通过 boxplot 图进行展示。我们通过使用 Spearman 算法分析了 TCGA-LUAD 数据集中不同免疫细胞之间的相关性, 并利用 R 包 ggplot2 进行了图形化展示。此外, 我们还探讨了免疫细胞与 MiAg DEGs 之间的相关性, 并使用 R 包 ggplot2 生成了相关性散点图以便于结果的可视化。

九、预后临床相关性分析

在 TCGA-LUAD 数据集中我们依据 MiAg Scores 高低分组构建多因素 Cox 模型。首先对 MiAg DEGs 预后模型的 MiAg Scores 和临床变量进行单因素 Cox 回归分析, 接着将所有变量引入多因素 Cox 回归分析中以构建多变量 Cox 模型。基于这些多因素 Cox 回归分析的结果, 我们使用 R 包 rms 来制作列线图 (即 nomogram 图), 旨在直观展示模型预测结果。我们使用的 nomogram 图 [Park et al, 2018] 是一种在直角坐标系上通过多条不相交的线段展示多个独立变量间函数关系的方法。此图基于多因素回归分析, 采用特定的标尺评分系统来表征模型中各个变量的重要性, 并据此计算总评分以预测事件的发生概率。

为了评价多因素 Cox 模型的准确性和分辨力, 我们应用了校准曲线图 (Calibration 图) [Perkins et al, 2019], 该图通过比较实际概率与模型预测概率的拟合程度来分析模型的预测效果。最后我们利用决策曲线分析 (Decision curve analysis,

DCA) [Van Calster et al, 2018], 这是一种简洁的方法, 用于评估临床预测模型、诊断试验和分子标记物的有效性, 通过 DCA 图评估模型的准确性和分辨力。我们用 R 包 ggDCA[Tataranni et al, 2019]绘制 DCA 图对多因素 Cox 模型的效果进行评估。预后模型构建的技术路线图如图 1 所示。

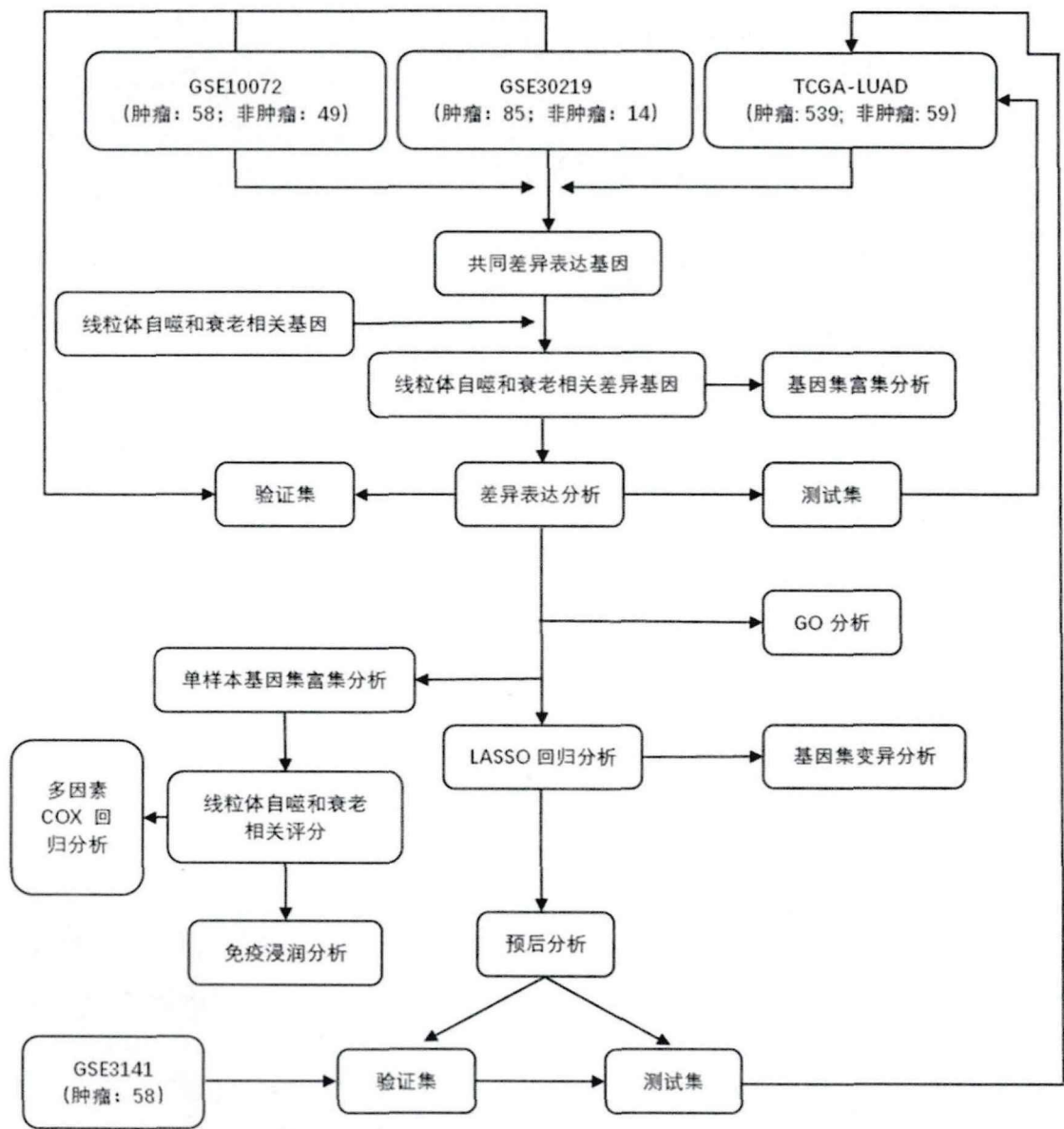


图 1. 线粒体自噬和衰老相关基因构建 LUAD 预后模型的技术路线图

十，入组 LUAD 患者选择

本论文中的相关组织取材和回顾性研究得到了中国医学科学院肿瘤医院伦理委员会的批准 (NCC2019C-167)。我们回顾性地收集 2018 年 1 月-12 月于我院胸外科手术的患者 110 例。患者的入组标准为: 1, 石蜡病理结果证实为肺腺癌; 2, 2018

年间在我院首次就诊并接受手术治疗；3，术后每年都有随访记录。排除标准为：1，未实施手术的患者；2，存在其他合并癌症或严重的基础疾病；3，基本临床资料不完整或丢失或失访；4，术后病理示病理类型为其他类型的肺癌。最终共 96 名患者（4 例术后病理合并其他类型的肺癌；8 例患者失访；2 例患者临床资料不完整）被纳入后续的分析用于多重荧光免疫组织化学实验的验证。

我们选取了 2022 年就诊于我院的 30 例 LUAD 患者的新鲜肺腺癌组织和癌旁组织各约 30mg 用于实时荧光定量聚合酶链式反应（Real Time Quantitative Polymerase Chain Reaction, Real Time-qPCR）。患者的入排标准同上。最终共 22 名患者（5 例术后病理合并其他类型的肺癌；3 例术前合并其他类型的癌症）纳入最终分析。

十一，Real-Time qPCR 实验

1，引物序列

基因名称	基因序列
CAV1-F	GCGACCCTAAACACCTCAAC
CAV1-R	ATGCCGTCAAAACTGTGTGTC
DSG2-F	CTAACAGGTTACGCTTTGGATGC
DSG2-R	GTGAACACTGGTTCGTTGTCAT
DSP-F	GCAGGATGTACTATTCTCGGC
DSP-R	CCTGGATGGTGTCTCTGGTTCT
MYH11-F	CGCCAAGAGACTCGTCTGG
MYH11-R	TCTTTCCCAACCGTGACCTTC
PLOD2-F	CATGGACACAGGATAATGGCTG
PLOD2-R	AGGGGTTGGTTGCTCAATAAAAA
β-actin-F	CCTGGCACCCAGCACAAT
β-actin-R	GGGCCGGACTCGTCATAC

2，实验步骤

（1）RNA 提取

- 1) 将 Trizol 裂解液转移到 EP 管中，室温(15-30℃)静置 5 分钟。
- 2) 向 EP 管中添加 0.2 mL 的氯仿，紧闭盖子并颠倒混匀 15 秒，在室温(15-30℃)下放置 2-3 分钟，之后在 2-8℃ 下以 12,000 g 的速度离心 15 分钟。
- 3) 轻取上层水相，转移到新的 EP 管中，按每 1 mL Trizol 添加 0.5 mL 异丙醇

的比例加入异丙醇，室温(15-30℃)放置 10 分钟后在 2-8℃ 下 12,000 g 离心 10 分钟。

4) 去除上清液，每 1 mL Trizol 加入 1 mL 75%乙醇进行洗涤，经涡旋混合后在 2-8℃ 下 7,500 g 离心 5 分钟，去除上清。

5) 允许 RNA 沉淀在室温下自然干燥后，用 Rnase-free 水溶解。

6) 使用紫外分光光度计和琼脂糖凝胶电泳检测 RNA 的纯度和完整性，并将其在-80℃ 超低温冰箱中保存。

(2) cDNA 模板的合成

1) 在无核酸酶的微量离心管中配置Real-Time qPCR反应体系-1 (20 μl)

组分	体积
5×gDNA Eraser Buffer	2 μl
gDNA Eraser	1 μl
总RNA	<1 μg
灭菌蒸馏水	补足至10 μl

轻轻混合后在42℃孵育2分钟或室温孵育两次5分钟。短暂离心后，加入反应体系-2 (20 μl) 组分：

组分	体积
步骤 1的反应液	10 μl
PrimeScript RT Enzyme Mix I	1
RT Primer Mix	1
5×PrimeScript Buffer 2	4
RNase Free dH ₂ O	4

2) 在 37 ℃孵育 15 min。

3) 在85℃加热5秒以终止反应。之后将cDNA产物稀释三倍，直接用作q-PCR的模板。

(3) 样本检测

反应所用 Real Time qPCR 引物的量根据不同 Real Time qPCR 试剂盒的说明进行调整，以 UE 公司 SYBR Green Mix(2X)定量 PCR 为例。

1) 按下列组分配制反应液。

试剂	使用量
SYBR Green Mix(2X)	10 μL
模板(cDNA 溶液)	1 μL
Real Time qPCR 正向引物(10μM) *	0.8 μL
Real Time qPCR 反向引物(10μM) *	0.8 μL

dH₂OUp to 20 μ L

2) Real Time qPCR 反应, 反应条件如下:

预变性 95°C 30 s, 再执行(95°C 10 s, 60°C 30 s, 72°C 30 s)共 40 个循环,

3) 实验结果分析。

十二, LUAD 中单细胞测序分析

1, 数据收集和信息整理

我们从 GEO[Barrett et al, 2013]和 ArrayExpress[Athar et al, 2019]中收集了肺腺癌 scRNA-seq 数据集 (EMTAB6149, GSE127465, 和 GSE143423)。下载了每个数据集的原始计数、TPM 或 FPKM 的表达矩阵。接着从数据库或原始研究中收集样本信息, 如患者 ID、组织来源和原始细胞类型注释。

2, 数据质量控制

我们应用基于 MAESTRO v1.1.0[Wang, C. et al, 2020]的标准化分析流程, 对所有收集到的数据集进行处理, 包括质量控制、批次效应去除、细胞聚类、差异表达分析、细胞类型注释和恶性细胞分类。原始计数、TPM 或 FPKM 表被用作标准化流程的输入。细胞质量由两个指标确定: 每个细胞的总计数 (UMI) 和每个细胞检测到的基因数量。如果库的大小小于 1000 或检测到的基因数小于 500, 则过滤掉低质量细胞。

3, 批次效应评估和校正

为了系统评估每个数据集的批次效应, 我们采用基于熵的指标[Azizi et al, 2018; Zhang et al, 2020]来衡量数据在批次间的混合程度。在大多数数据集中, 来自不同患者的样本通常受到批次效应的影响。我们对每个具有多个患者的数据集, 基于 UMAP 坐标系中细胞之间的欧氏距离构建了一个 k-NN 图。然而, 需要注意的是, 对于主要包含恶性细胞的数据集, 低熵可能来自于恶性细胞表达异质性的存在[Puram et al, 2017]。批次效应也应用 Seurat v3.1.2 方法进行了去除。

4, 细胞聚类和差异基因分析

对于每个数据集 (EMTAB6149, GSE127465, 和 GSE143423), MAESTRO 流程识别了前 2000 个可变特征, 并使用 PCA 进行降维, 使用 KNN 和 Louvain 算法进行聚类识别[Stuart et al, 2019; Xu et al, 2015]。为了更好地捕捉具有不同细胞数量的数据集的细胞差异和变异性, 我们根据细胞数量增加了主成分的数量和用于基于图的聚类的分辨率。采用 UMAP 进一步降低维度并可视化聚类结果[Becht et al, 2018]。我们采用 Wilcoxon 检验, 根据对数转换的折叠差异 ($|\log FC| \geq 0.25$) 和错误发现率 ($FDR < 1e-05$), 识别每个聚类与所有其他细胞之间的差异表达基因。

5, 细胞类型注释

通过结合三种方法确定恶性细胞的聚类。首先,我们采用原始研究提供的细胞类型注释。其次,我们根据初始研究中的恶性细胞标记物表达分布,检查恶性细胞的特征[Lambrechts et al, 2018]。第三,在预测的拷贝数变异的基础上,我们运行了 InferCNV (v1.2.1)[Patel et al, 2014]来预测细胞的恶性程度,并将细胞分成恶性细胞和非恶性细胞的聚类。对于其他正常的聚类,我们采用 MAESTRO 中基于标记基因的注释方法,使用聚类间的差异表达基因进行自动注释。每个细胞类型的标记基因来自自己发表的资源[Newman et al, 2015; Yost et al, 2019; Zhang et al, 2018],并进行手动整理。我们计算了每个聚类中每个细胞类型的标记基因的平均 logFC,并将其作为细胞类型评分。每个聚类将被分配一个具有在所有输入细胞类型标记中最高评分的特定细胞类型。

十三, 多重荧光免疫组织化学

1. 脱蜡至水:将石蜡切片依次放入三种环保型脱蜡液(G1128)中各 10 分钟,然后依次通过三级无水乙醇各 5 分钟,最后用蒸馏水清洗。
2. 抗原修复:使用 pH 6.0 的柠檬酸修复液进行高压加热,待喷气 2 分钟后停止加热,并用冷自来水冲洗直到冷却。注意防止修复液过度蒸发和干燥片子。冷却后,将玻片在 pH 7.4 的 PBS 中于钟摆摇床上洗涤 3 次,每次 5 分钟。
3. 封闭处理:使用免疫组化笔在切片周围画圈,防止试剂外流,然后在避光湿盒中滴加 3%双氧水溶液,室温避光孵育约 25 分钟,封闭内源性过氧化物酶。之后同样在 PBS 中洗涤。
4. 血清封闭:切片放于避光湿盒中,滴加适宜的血清进行封闭,如使用山羊抗体的样本用 10%兔血清封闭,其他样本用 3% BSA 封闭,室温孵育 30 分钟。
5. 一抗应用:轻轻甩掉封闭液,向切片上滴加稀释好的一抗溶液,放入避光湿盒中 4°C 孵育过夜。
6. 二抗应用:切片在 PBS 中洗涤后,轻轻甩干,然后在已画好的圈内滴加对应种属的 HRP 标记二抗,室温孵育 50 分钟。
7. 信号放大:应用 TSA 增强信号,处理后在 PBS 中洗涤,然后在避光条件下室温孵育 10 分钟。
8. 微波处理:将组织切片放入含柠檬酸修复液的盒子中,在微波炉中加热约 10 分钟,移除已结合的抗体,注意防止溶液过度蒸发。
9. 图像采集:按照设定的荧光素颜色及激发与发射波长进行图像采集,推荐使用专业设备进行全景扫描。

十四, 细胞培养及转染

1, 细胞系

人源肺癌细胞系 H157、A549、H1395 和 H1975 以及正常人支气管上皮细胞 BEAS-2B 均购自武汉普诺赛生命科技有限公司。

2, 细胞转染

(1) 细胞处理:

由北京欧林格生物科技有限公司负责 DSP 过表达质粒构建。NC-siRNA 和 DSP-siRNA (5'-AGAGCUCAAUAAGCUGAAACAAGAA-3') 也由北京欧林格生物科技有限公司合成。

(2) 转染步骤:

- 1) 转染前一天, 将 2×10^5 密度细胞接种在 6 孔板上, 加入 2mL 含 FBS 的 DMEM 细胞培养基;
- 2) 在 250 μ l Opti-MEM 培养基中加入 100 uM siRNA, 柔和混匀, 室温孵育 5min;
- 3) 混匀 SiMi Transfection Reagents 试剂, 用 250 μ l 的 Opti-MEM 的培养基稀释 10 μ l SiMi Transfection Reagents 试剂, 柔和混匀, 室温孵育 5min;
- 4) 将稀释好的 siRNA 和 SiMi Transfection Reagents 试剂混合;轻柔混匀, 室温孵育 20min,以便形成 siRNA/SiMi Transfection Reagents 复合物。
- 4) 将 500 μ l siRNA/SiMi Transfection Reagents 复合物加到含有细胞和培养基的培养板的孔中, 来回轻柔摇晃细胞培养板。
- 6) 转染 4-6 小时后, 除去复合物, 更换新鲜完全培养基, 继续培养 48-72h, 用于后续检测。

十五, 蛋白免疫印记 (Western Blot) 检测

1, 蛋白提取

- (1) 添加 1 mL PBS, 500g 离心 5 分钟。
- (2) 吸去残余 PBS, 估计样品体积, 添加 5 倍体积的裂解液, 充分混匀。
- (3) 在冰浴中使用 20%功率进行三次 10 秒的超声破碎处理。
- (4) 在 4°C 和 12,000 rpm 条件下离心 15 分钟, 收集上清液。
- (5) 使用 BCA 法测定蛋白浓度, 调整至 4 mg/mL, 并在 -20°C 保存, 防止反复冻融。

2, 蛋白浓度测定

- (1) 准备 BSA 标准品稀释液。
- (2) 配制 BCA 工作液。
- (3) 将标准品和样品蛋白加入到孔板或试管中, 加入 BCA 工作液, 混匀, 并在 37°C 保温 30-60 分钟。

(4) 冷却至室温后, 测量 562nm 或其附近波长的吸光值, 并减去空白对照的平均吸光值。

3, SDS-PAGE 电泳

(1) 准备分离胶和浓缩胶, 确保在聚合过程中不形成气泡, 并在聚合过程中用去离子水覆盖凝胶。

(2) 凝胶聚合后, 将其装入电泳装置中, 加载蛋白样品。

(3) 进行电泳, 直到染料到达分离胶的适当位置。

4, 蛋白转移和 Western Blot

(1) 将 PVDF 膜先用甲醇浸泡, 然后用水和转移缓冲液漂洗。

(2) 组装转移并在恒定电压下进行转移。

(3) 封闭膜, 先后用一抗和二抗孵育, 清洗, 用化学发光检测法进行显影。

十六, CCK8 细胞增殖实验

1, 以每孔 1×10^5 个细胞量/100uL 接种到 96 孔板中, 对照孔加入等量的培养基, 37°C, 5% CO₂ 细胞培养箱培养;

2, 细胞贴壁后, 分别于相应的时间点每孔加入 10uL CCK-8 溶液, 培养箱中继续培养 1 小时;

3, 酶标仪测定 450nm 出的吸光度。

十七, 平板克隆形成实验

1, 实验前准备: 6 孔板、吸管、枪头、血球计数板等放入超净工作台消毒 30min。

2, 实验流程:

(1) 取出贴壁率>90%的细胞(处于对数生长期), 消化离心, 重悬细胞并进行细胞计数。

(2) 6 孔板种板: 实验分为空白对照、DSG 过表达组, NC-siRNA 和 DSG2-siRNA 组, 每个孔中加入 1000 个细胞, 放入 CO₂ 培养箱培养。

(3) 培养 1 周后, 当肉眼可见集落形成时, 取出 6 孔板, 弃除培养液, 每孔加入 2mL 甲醇固定 30min, 弃除甲醇, 每孔加入 2mL 0.1% 结晶紫染色 3min, 然后清洗掉结晶紫, 数码相机拍照, 光镜下计数集落数, 统计数据。

十八, Transwell 实验

1, 细胞样本的准备与处理

(1) 细胞培养：在实验开始前，将细胞在不含血清的培养基中进行 12 小时的饥饿处理。

(2) 细胞处理：应用胰蛋白酶对细胞进行消化处理。消化结束后，以 1000 转/分钟的速度离心 3 分钟，去除上层清液。随后，利用磷酸盐缓冲液 (PBS) 洗涤细胞并去除残留上清，再次用不含血清的培养基 (添加 0.1% 牛血清白蛋白 BSA) 重新悬浮细胞，并进行数量计数。

(3) 细胞密度调整：使用不含血清的 0.1% BSA 培养基调整细胞浓度至每毫升 1×10^5 细胞。

(4) 培养板准备：在一个 24 孔的培养板中加入 700 微升的完全培养基。

(5) Transwell 装置准备：在 Transwell 装置的上室内加入 200 微升的细胞悬液。

(6) 转移与培养：使用镊子将装置转移至含有完全培养基的 24 孔板中，注意避免气泡产生。然后，将培养板放入设定为 37°C、5% CO₂ 的培养箱中，培养 24 至 48 小时。

2, 细胞染色过程

(1) 移除培养基与细胞清洁：轻轻移除 Transwell 上室的培养基，并用 PBS 润湿的棉签清除上室内部的细胞残留。

(2) 细胞固定与染色：将 Transwell 的底面浸泡在 10% 的甲醇中 30 秒以固定细胞，随后转移到纯净水中以去除甲醇。接着将底面浸入结晶紫染液中进行 2 分钟的染色，之后用纯净水清洗直至背景变得清晰。

3, 成像与计数

在 100 倍光学显微镜下随机选择五个视野，计算穿过膜的细胞数量。

十九, 划痕实验

1, 标记线条：在 6 孔板的背面使用标记笔，借助直尺画出间隔为 0.5 至 1 厘米的横线，确保每个孔穿过至少五条线。

2, 细胞接种：在对数生长期的细胞中使用胰蛋白酶处理至单细胞悬液状态，将细胞接种到每个孔的 6 孔培养板中，接种量为 6 万细胞/孔，以确保过夜后细胞完全覆盖孔底，每孔添加 2 毫升培养基。

3, 细胞培养：将培养板放入设置为 37°C、5% CO₂ 的培养箱中，持续培养 24 小时。

4, 制作划痕：第二天，使用 200 微升的吸头，参照 6 孔板的盖子或直尺画线，确保划线平行或垂直于板背的线条，并保持吸头垂直，避免倾斜。

5, 细胞清洗：使用 PBS 溶液将细胞表面冲洗三次，移除因划痕而松动的细胞，随后加入无血清培养基。

6, 影像记录：清除 6 孔板背面的标记线。使用 4 倍放大镜进行拍照，确保每个划痕

处于视野中心且垂直，背景保持一致。在进行样本测试时，设置不同浓度梯度，并在 0、6、12、24 小时各时间点进行拍照记录。

二十，流式细胞学技术

1，细胞样本制备：

(1) 细胞收集：将培养液转移到离心管中保存。利用不含 EDTA 的胰蛋白酶处理细胞，直至细胞能轻易通过移液管吹打脱落。将先前的培养液加回，吹打脱落所有贴壁细胞，并轻柔分散。之后再次将细胞收集至离心管中。

(2) 离心处理：在大约 1000 转/分钟的速度下离心 5 分钟，沉积细胞。针对难以沉积的细胞，适当增加离心时间或力度以确保细胞完全沉积。

(3) 上清处理：轻轻移除上层清液，留下大约 50 微升培养液以避免吸取细胞。

(4) PBS 清洗：加入大约 1 毫升 4°C 的 PBS 重悬细胞，再次离心并移除清液。

(5) 缓冲液处理：将 4 毫升 4 倍结合缓冲液与 12 毫升去离子水按 1:3 的比例稀释。

(6) 细胞悬浮：使用 1 倍结合缓冲液重悬细胞，并调整细胞浓度至 1 至 5×10^6 /毫升。

2，流式细胞检测：

(1) 样品准备：取 100 微升的细胞悬液放入 5 毫升的流式管中，添加 5 微升的 Annexin V/FITC 并在室温下避光孵化 5 分钟。

(2) 染色与检测：向样品中添加 10 微升 20 微克/毫升的碘化丙啶溶液和 400 微升 PBS，随后立即进行流式细胞检测。

二十一，统计分析

本篇文章的所有数据处理与分析皆基于 R 软件 (Version 4.2.2)，对于两组连续数据的比较，正态分布的数据通过独立的 Student's t 检验来评估其统计学意义。对于非正态分布的数据，则采用 Mann-Whitney U 检验，也称为 Wilcoxon 秩和检验，来进行差异性分析。此外，当涉及到两组分类数据的比较时，使用卡方检验或 Fisher 的精确检验来确定其统计显著性。R 的 survival 包被用于进行生存分析，Kaplan-Meier 生存曲线用于显示生存差异，对数秩检验 (Log-rank test) 用于评估两组患者生存时间差异的显著性。单因素和多因素 Cox 分析均基于 survival R 包。所有的统计 P 值均为双侧，以 $P < 0.05$ 为有统计学意义。

实验结果

一、LUAD 数据集的标准化处理

首先,我们对 TCGA-LUAD 数据集使用 R 包 limma 包进行标准化处理 (图 2A-B), TCGA-LUAD 数据集共有 598 例样本:包括 59 例正常组样本 (分组: Normal) 和 539 例癌症组样本 (分组: Tumor) (图 2A-B)。由图 2A-B 可知, TCGA-LUAD 数据集的表达谱数据在进行标准化处理后样本间的数据表达情况趋于一致。然后我们对 TCGA-LUAD 数据集的表达谱数据根据 Tumor 组和 Normal 组的分组情况,对去除批次效应前后数据集表达矩阵进行主成分分析 (Principal Component Analysis, PCA)。然后将 TCGA-LUAD 数据集去除批次效应前后的主成分分析 (PCA) 结果展示出来 (图 2C-D)。

然后我们对 GSE10072 数据集和 GSE30219 数据集使用 R 包 limma 包进行标准化处理 (图 3A-D)。GSE10072 数据集共有 107 例样本:49 例正常组样本 (分组: Normal) 和 58 例癌症组样本 (分组: Tumor) (图 2C-D); GSE30219 数据集共有 99 例样本:14 例正常组样本 (分组: Normal) 和 85 例癌症组样本 (分组: Tumor) (图 2E-F)。由图 3A-D 可知, GSE10072 和 GSE30219 的表达谱数据在进行标准化处理后样本间的数据表达情况趋于一致。

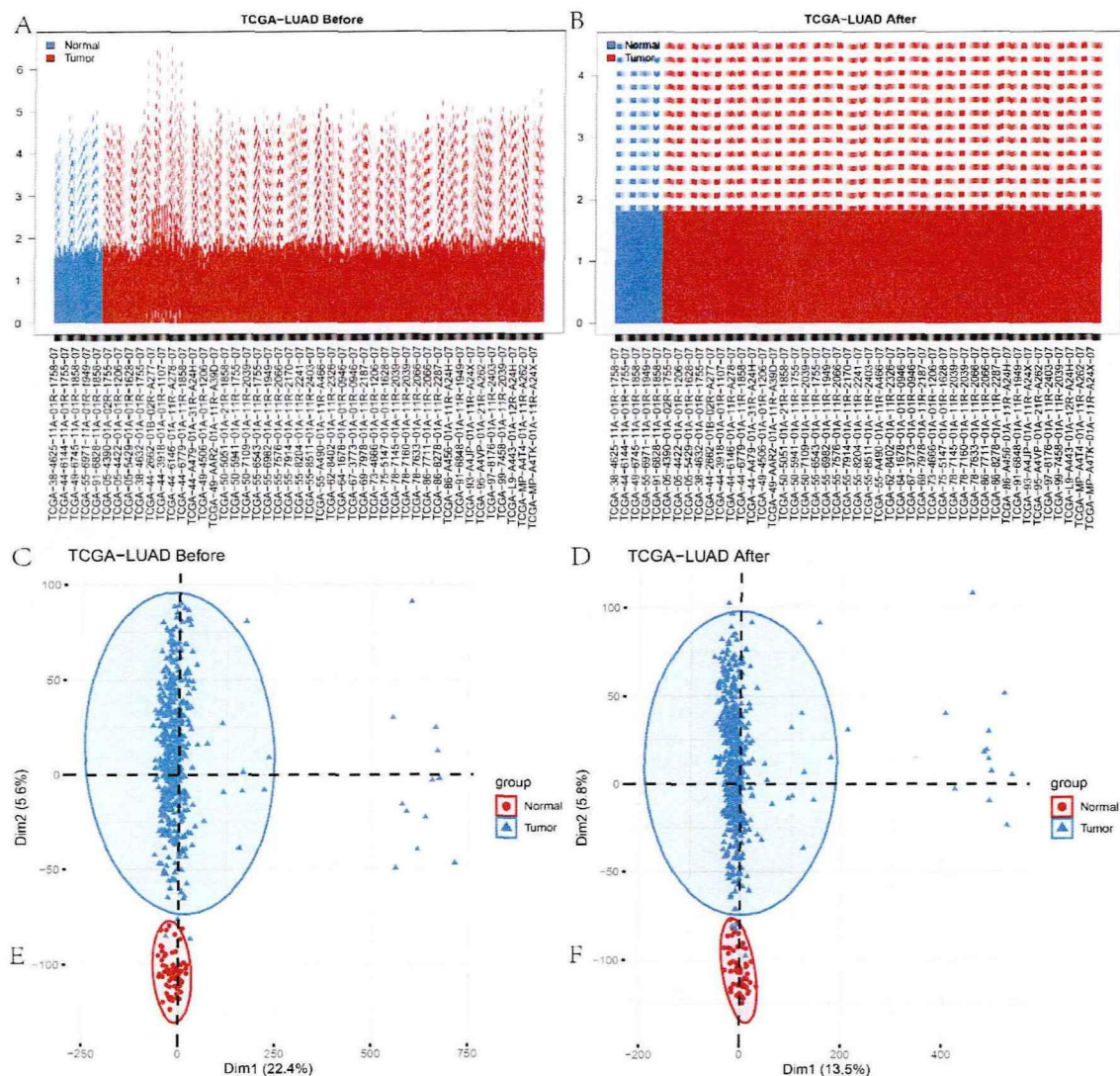


图 2. TCGA-LUAD 数据集的标准化处理

A-B. TCGA-LUAD 数据集标准化处理前 (A), 后 (B) 的 boxplot 图。C-D. TCGA-LUAD 数据集标准化处理前 (C), 后 (D) 的 PCA 图。

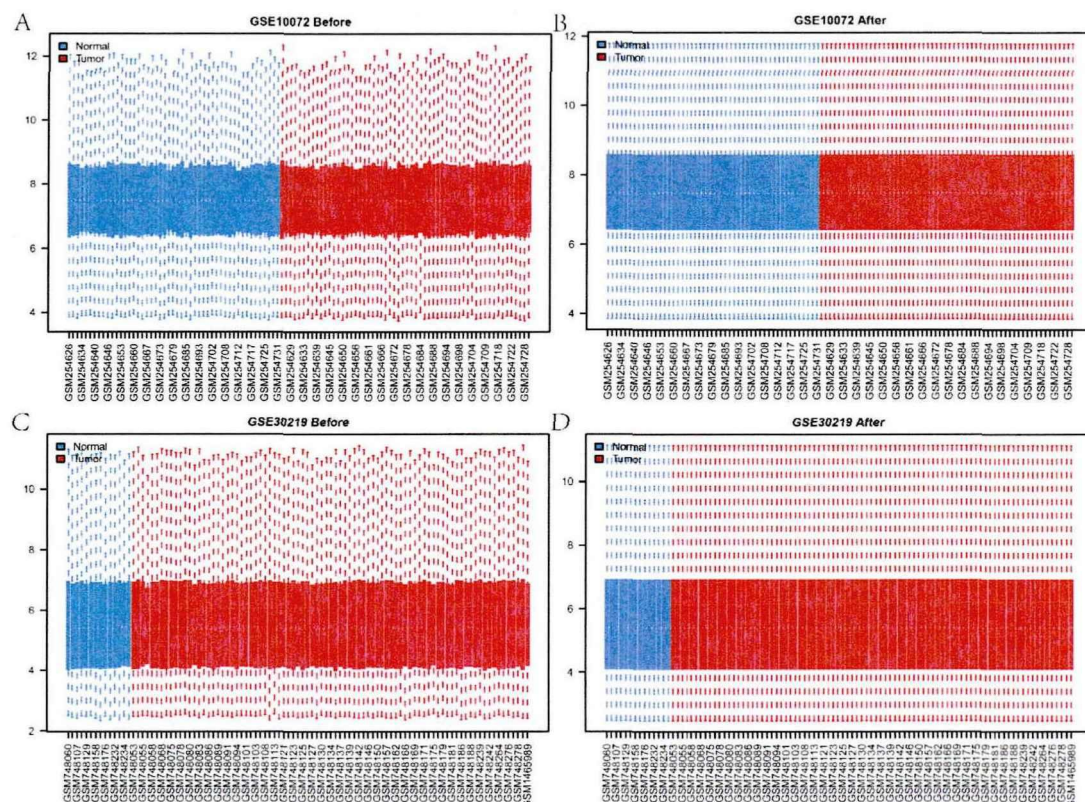


图 3. GEO 中 LUAD 数据集的标准化处理

A-B. GSE10072 数据集标准化处理前 (A), 后 (B) 的 boxplot 图。C-D. GSE30219 数据集标准化处理前 (C), 后 (D) 的 boxplot 图。

二, 线粒体自噬和衰老相关基因在 LUAD 患者中的突变分析

为了分析 560 个线粒体自噬和衰老相关基因在 TCGA-LUAD 数据集中 LUAD 患者的体细胞突变情况, 我们统计了 TCGA-LUAD 数据集 LUAD 患者样本中 560 个线粒体自噬和衰老相关基因的突变分析, 并通过 R 包 *maftools* 包对结果进行可视化。结果显示: TCGA-LUAD 数据集里体细胞突变的主要种类共有 7 种, 分别是: 错义突变、无义突变、移码缺失突变、剪接位点、移码插入突变、框内缺失、框内插入、终止子突变, 以错义突变为主 (图 4A)。

另外, 560 个线粒体自噬和衰老相关基因在 LUAD 患者中的突变类型主要是单核苷酸多态性 (Single Nucleotide Polymorphism, SNP), 以及少量的插入和缺失。而且 C > T 在 LUAD 患者的单核苷酸变体 (Single Nucleotide Variations, SNV) 中最常见, 其次是 T > C, C > A 等 (图 4A)。图 4A 表明, TCGA-LUAD 数据集里 560 个线粒体自噬和衰老相关基因中存在体细胞突变最多的基因是 TP53, 涉及 6 种突变类型。另外在 TCGA-LUAD 数据集中存在体细胞突变最多的 10 个线粒体自噬和

衰老相关基因分别是 TP53, KRAS, KEAP1, SYNE2, NLRP3, PLEC, VPS13C, HUWE1, ATM, PRKDC, 涉及最多的突变类型为错义突变。然后我们还将 LUAD 患者中的体细胞突变最多的 20 个线粒体自噬和衰老相关基因全部展示出来(图 4B)。TCGA-LUAD 数据集中体细胞突变最多的 20 个线粒体自噬和衰老相关基因的相关性结果显示(图 4C), TCGA-LUAD 数据集里体细胞突变最多的 20 个线粒体自噬和衰老相关基因的体细胞突变情况中相当一部分线粒体自噬和衰老相关基因存在显著相关性 ($P<0.05$)。

通过分析 MiAg RGs 在 TCGA-LUAD 数据集中的拷贝数变异 (Copy Number Variations, CNV) 情况 (展示拷贝数改变频率最高的前 20 个基因) (图 4D), 发现大部分样本均存在拷贝数改变情况, 且大多数样本集中在拷贝数缺失上, 如 ATP1B1、CCT3、CD55 及 DAP3 等基因具有较高的拷贝数扩增频率 (40%-60%), 但几乎没有出现拷贝数缺失 (图 4E)。部分基因在 LUAD 患者中具有更高的拷贝数改变频率 (图 4D), 如 FDPS、ILF2 和 MLC1 等。

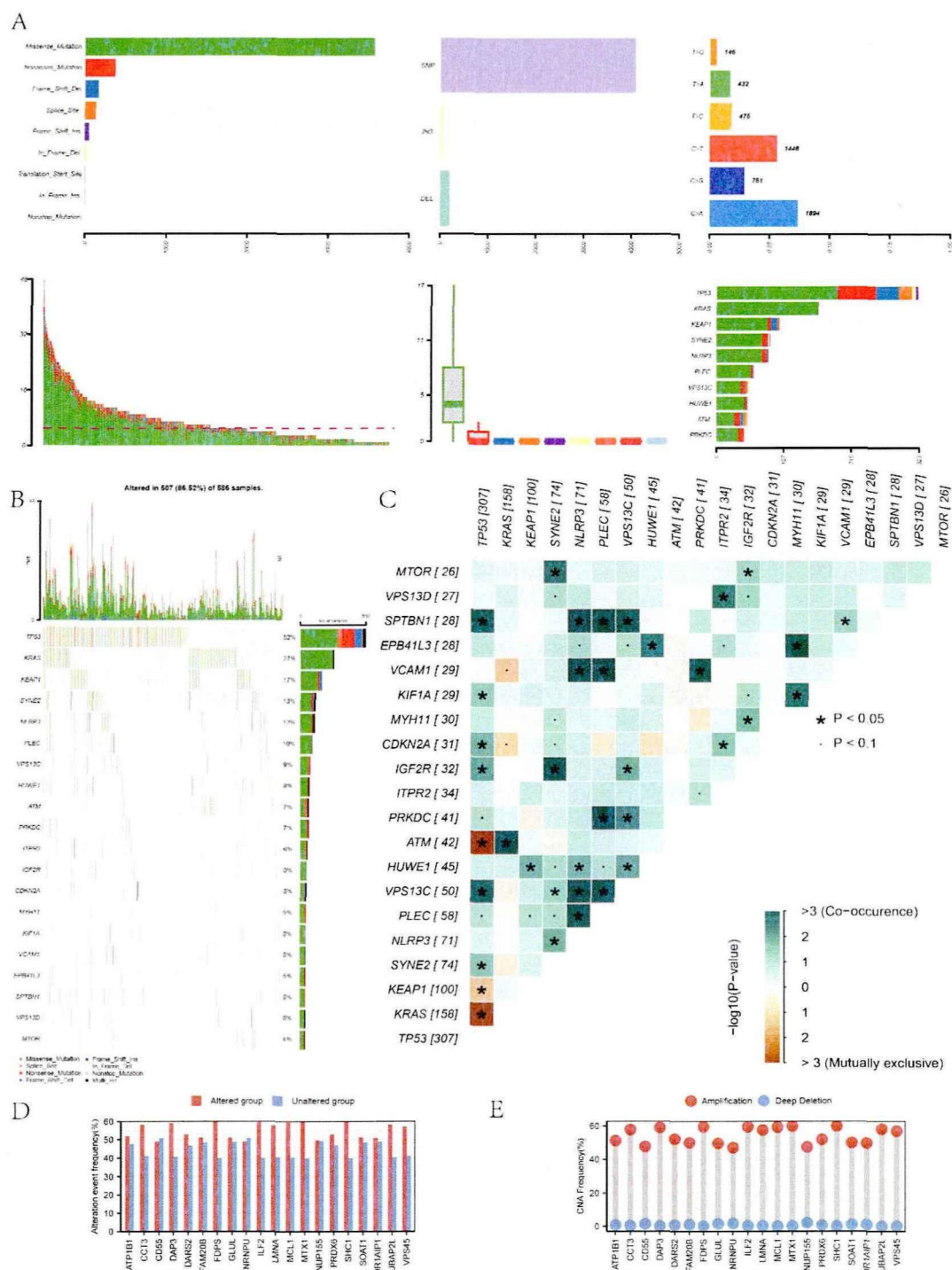


图 4. 线粒体自噬和衰老相关基因在肺腺癌患者中的突变分析

A. 线粒体自噬和衰老相关基因在 LUAD 中的 SNP 情况展示。B-C. 线粒体自噬和衰老相关基因在 LUAD 中的体细胞突变情况占比结果展示 (B), 体细胞突变相关性分析结果 (C) 展示。D-E. 线粒体自噬和衰老相关基因在 LUAD 中的 CNV 改变频率 (D) 和改变类型 (E) 展示。

三, LUAD 相关差异基因分析

为了分析基因在 LUAD 肿瘤组织(分组:Tumor)相对于正常组织(分组:Normal)的表达差异,通过使用 limma 软件包,我们对 TCGA-LUAD、GSE10072 和 GSE30219 数据集进行了基因表达差异性分析。在这些肺腺癌数据集中,我们识别出了多个差异表达基因。具体来说,TCGA-LUAD 数据集中,共识别出 56,683 个差异表达基因。在这些基因中,根据 $|\log FC| > 1$ 和 $P_{adj} < 0.05$ 的标准,有 7,413 个基因显示出显著差异,其中 4,095 个基因在肿瘤样本中表达上调,3,318 个基因表达下调。差异表达的结果通过火山图(见图 5A)直观展示;在 GSE10072 数据集中,共识别出了 12,548 个差异表达基因。遵循 $|\log FC| > 1$ 和 $P_{adj} < 0.05$ 的筛选标准,447 个基因表现出显著的表达差异。具体地,299 个基因表达水平上调,148 个基因表达下调。我们利用这些数据绘制了火山图(见图 5B),以直观展示基因表达的变化;在 GSE30219 数据集中,发现了总共 21,655 个差异表达基因。按照 $|\log FC| > 1$ 且 $P_{adj} < 0.05$ 的标准,筛选出 1,037 个基因显示显著的表达差异。其中,697 个基因呈现上调,340 个基因呈现下调。关于这些发现,我们制作了火山图(见图 5C)以直观显示基因表达变化。

为了得到线粒体自噬和衰老相关差异表达基因 (MiAg related differentially expressed genes, MiAg DEGs),我们先对 TCGA-LUAD 数据集, GSE10072 数据集和 GSE30219 数据集中得到的全部 $|\log FC| > 1$ 且 $P_{adj} < 0.05$ 的差异表达基因取交集,得到了肺腺癌数据集的共同差异表达基因 300 个(图 5D)。我们又对数据集中的共同差异表达基因和线粒体自噬和衰老相关基因取交集,共得到 7 个肺腺癌的 MiAg DEGs 并绘制了韦恩图(图 5E),这 7 个 MiAg DEGs 分别是: CAV1, DSG2, DSP, MYH11, NME1, PAICS, PLOD2。

我们进一步分别对这 7 个 MiAg DEGs 在 TCGA-LUAD 数据集(图 5F),我们对 GSE10072(图 5G)和 GSE30219(图 5H)数据集中肿瘤与正常组的表达差异进行了详细分析。通过使用 R 语言的 pheatmap 包,我们生成了热图以展示这些差异。(图 5F-J)。

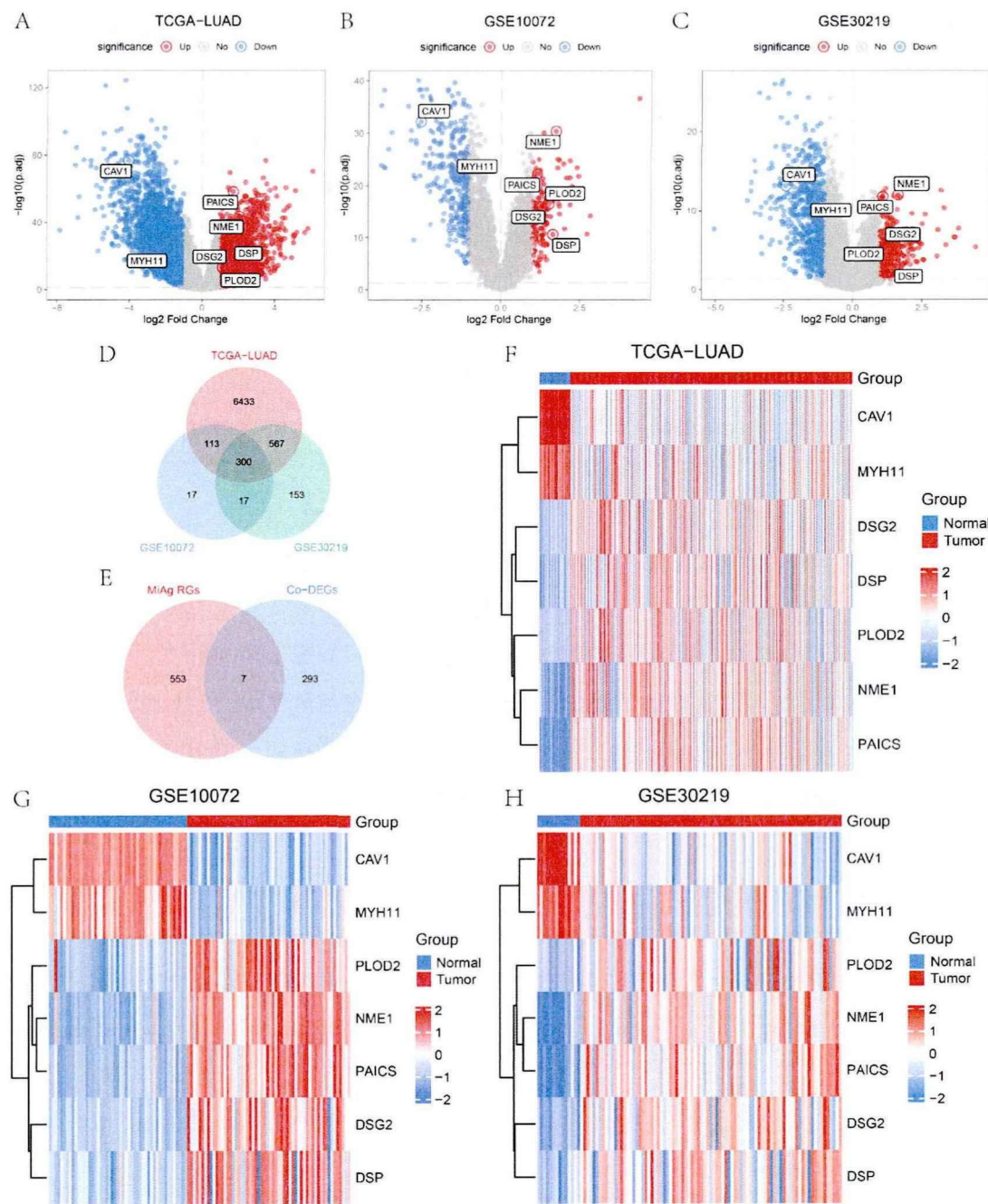


图 5. LUAD 中差异基因分析与线粒体自噬和衰老相关差异基因热图

A-C. 在 TCGA-LUAD 数据集 (A)、GSE10072 数据集 (B) 以及 GSE30219 数据集 (C) 中，我们展示了癌症组 (Tumor) 与正常组 (Normal) 之间差异表达基因的火山图分析。D. 展示了 TCGA-LUAD、GSE10072 和 GSE30219 数据集中差异表达基因的维恩图分析。E. TCGA-LUAD 数据集，GSE10072 数据集和 GSE30219 数据集 中的共同差异基因和线粒体自噬和衰老相关基因的韦恩图。F-H. TCGA-LUAD 数据 集 (F)，GSE10072 数据集 (G) 和 GSE30219 数据集 (J) 中 MiAg DEGs 的复杂数

值热图。

四, MiAg DEGs 的相关性分析

我们对 7 个 MiAg DEGs(CAV1, DSG2, DSP, MYH11, NME1, PAICS, PLOD2) 在 TCGA-LUAD 数据集, GSE10072 数据集和 GSE30219 数据集表达量之间的相关性进行了分析, 我们对 7 个 MiAg DEGs 在 TCGA-LUAD 数据集中的表达量进行 Spearman 相关性分析并以热图展示分析结果(图 6A)。结果显示在 TCGA-LUAD 数据集中 7 个 MiAg DEGs 的表达量之间都呈现出显著的相关性 ($P<0.05$)。我们对 7 个 MiAg DEGs 在 GSE10072 数据集中的表达量进行 Spearman 相关性分析并将结果通过热图进行展示(图 6D), 在 GSE10072 数据集中 7 个 MiAg DEGs 的表达量之间都呈现出显著的相关性 ($P<0.05$)。然后我们对 7 个 MiAg DEGs 在 GSE30219 数据集中的表达量进行 Spearman 相关性分析并进行热图展示(图 7A), 在 GSE30219 数据集中 7 个 MiAg DEGs 的表达量之间都呈现出显著的相关性 ($P<0.05$)。

我们还分别将相关性热图中相关性趋势相同的基因之间的相关性分析结果通过相关性散点图进行展示(图 6B-C, 图 6E-F, 图 7B-C), 7 个 MiAg DEGs 中 MYH11 和 NME1 ($r=-0.585$, $P<0.001$, 图 6B) 在 TCGA-LUAD 数据集中的表达量间表现出中等负相关性。NME1 和 PAICS ($r=0.616$, $P<0.001$, 图 6C) 在 TCGA-LUAD 数据集中的表达量间表现出中等正相关性; 而 MYH11 和 NME1 ($r=-0.775$, $P<0.001$, 图 6E) 在 GSE10072 数据集中的表达量间表现出中等负相关性, NME1 和 PAICS ($r=0.860$, $P<0.001$, 图 6F) 在 GSE10072 数据集中的表达量间表现出强正相关性; 而 MYH11 和 NME1 ($r=-0.722$, $P<0.001$, 图 7B) 在 GSE30219 数据集中的表达量间表现出中等负相关性, NME1 和 PAICS ($r=0.749$, $P<0.001$, 图 7C) 在 GSE30219 数据集中的表达量间表现出中等正相关性。(相关性散点图中的相关系数 r 绝对值在 0.8 以上为强相关; 绝对值在 0.5-0.8 为中等程度相关; 绝对值在 0.3-0.5 为弱相关; 绝对值在 0.3 以下为弱或者不相关。)

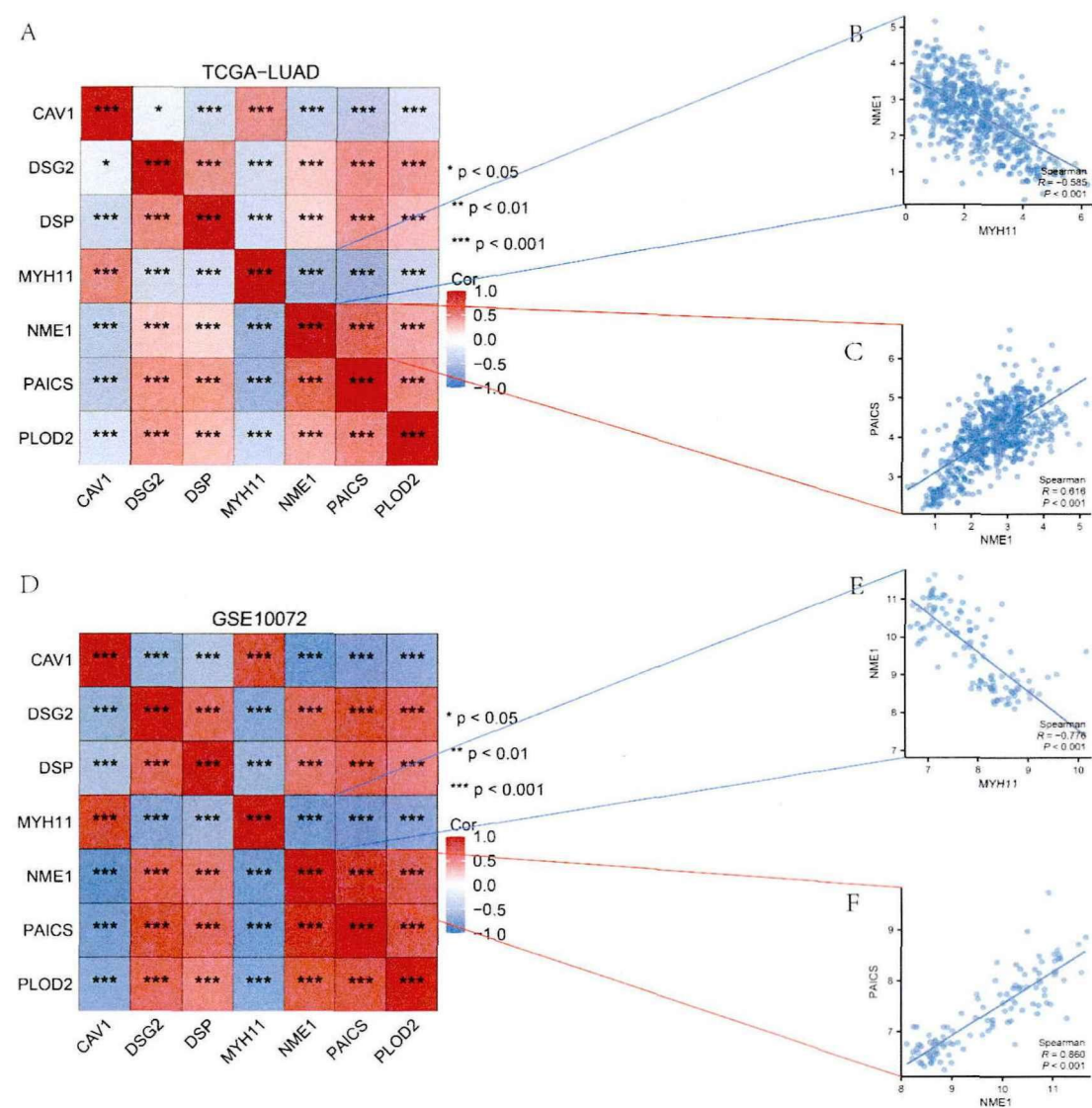


图 6. TCGA-LUAD 和 GSE10072 数据集中 MiAg DEGs 的相关性分析

A. MiAg DEGs 在 TCGA-LUAD 数据集中相关性热图结果展示。B-C. MiAg DEGs MYH11 和 NME1 (E), NME1 和 PAICS (F) (C) 的相关性散点图结果展示。D. MiAg DEGs 在 GSE10072 数据集中相关性热图结果展示。E-F. MiAg DEGs MYH11 和 NME1 (E), NME1 和 PAICS (F) 的相关性散点图结果展示。(*, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$)

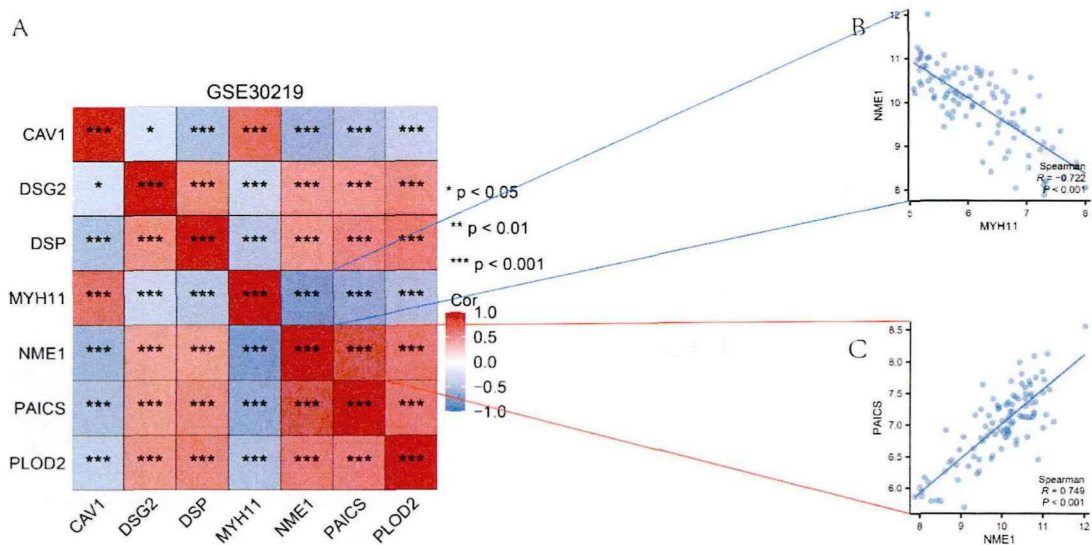


图 7. GSE30219 数据集中 MiAg DEGs 的相关性分析

A. MiAg DEGs 在 GSE30219 数据集中相关性热图结果展示。B-C. MiAg DEGs MYH11 和 NME1 (B), NME1 和 PAICS (C) 的相关性散点图结果展示。(*, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$)

五、MiAg DEGs 的表达差异分析

为深入探究 7 个 MiAg DEGs (CAV1, DSG2, DSP, MYH11, NME1, PAICS, PLOD2)在三个 LUAD 数据集(TCGA-LUAD 数据集,GSE10072 数据集和 GSE30219 数据集)中 LUAD 患者癌症组(分组: Tumor)和正常组(分组: Normal)的表达差异,我们对这些 MiAg DEGs 在 TCGA-LUAD 数据集, GSE10072 数据集和 GSE30219 数据集中的表达量高低与 Tumor 组和 Normal 组之间的相关性进行进一步的分析。

我们首先通过 Wilcoxon 符号秩和检验对 7 个 MiAg DEGs(CAV1, DSG2, DSP, MYH11, NME1, PAICS, PLOD2)在 TCGA-LUAD 数据集中 Tumor 组和 Normal 组间的表达量进行表达差异分析,结果显示 7 个 MiAg DEGs (CAV1, DSG2, DSP, MYH11, NME1, PAICS, PLOD2)在 TCGA-LUAD 数据集中 Tumor 组和 Normal 组间的表达差异都具有显著的统计学意义(图 8A)。然后通过 Wilcoxon 符号秩和检验对 7 个 MiAg DEGs (CAV1, DSG2, DSP, MYH11, NME1, PAICS, PLOD2)在 GSE10072 数据集中 Tumor 组和 Normal 组间的表达量进行表达差异分析,结果显示 7 个 MiAg DEGs (CAV1, DSG2, DSP, MYH11, NME1, PAICS, PLOD2)在 GSE10072 数据集中 Tumor 组和 Normal 组间的表达差异都具有显著的统计学意义(图 9A)。然后通过 Wilcoxon 符号秩和检验对 7 个 MiAg DEGs (CAV1,

DSG2, DSP, MYH11, NME1, PAICS, PLOD2) 在 GSE30219 数据集中 Tumor 组和 Normal 组间的表达量进行表达差异分析, 结果显示 7 个 MiAg DEGs (CAV1, DSG2, DSP, MYH11, NME1, PAICS, PLOD2) 在 GSE30219 数据集中癌症组 (Tumor) 和正常组 (Normal) 间的表达差异都有显著的统计学意义 (图 10A)。

随后我们绘制了 7 个 MiAg DEGs (CAV1, DSG2, DSP, MYH11, NME1, PAICS, PLOD2) 在 TCGA-LUAD 数据集, GSE10072 数据集和 GSE30219 数据集的 ROC (Receiver operating characteristic curve) 曲线 (图 8B-H, 图 8B-H, 图 8B-H)。在 TCGA-LUAD 数据集中, 7 个 MiAg DEGs (图 8B-H): CAV1 (AUC=0.994, 图 8B), DSG2 (AUC=0.853, 图 8C), DSP (AUC=0.843, 图 8D), MYH11 (AUC=0.912, 图 8E), NME1 (AUC=0.939, 图 8F), PAICS (AUC=0.984, 图 8G), PLOD2 (AUC=0.812, 图 8H) 的表达与肺腺癌的发生呈现出显著相关性。在 GSE10072 数据集中, 7 个 MiAg DEGs (图 9B-H): CAV1 (AUC=0.993, 图 9B), DSG2 (AUC=0.947, 图 9C), DSP (AUC=0.864, 图 9D), MYH11 (AUC=0.968, 图 9E), NME1 (AUC=0.984, 图 9F), PAICS (AUC=0.984, 图 9G), PLOD2 (AUC=0.976, 图 9H) 的表达与 LUAD 的发生呈现出显著相关性。在 GSE30219 数据集中, 7 个 MiAg DEG (图 10B-H) s: CAV1 (AUC=0.948, 图 10B), DSG2 (AUC=0.915, 图 10C), DSP (AUC=0.834, 图 10D), MYH11 (AUC=0.950, 图 10E), NME1 (AUC=0.963, 图 10F), PAICS (AUC=0.982, 图 10G), PLOD2 (AUC=0.853, 图 10H) 的表达与肺腺癌的发生呈现出显著相关性。

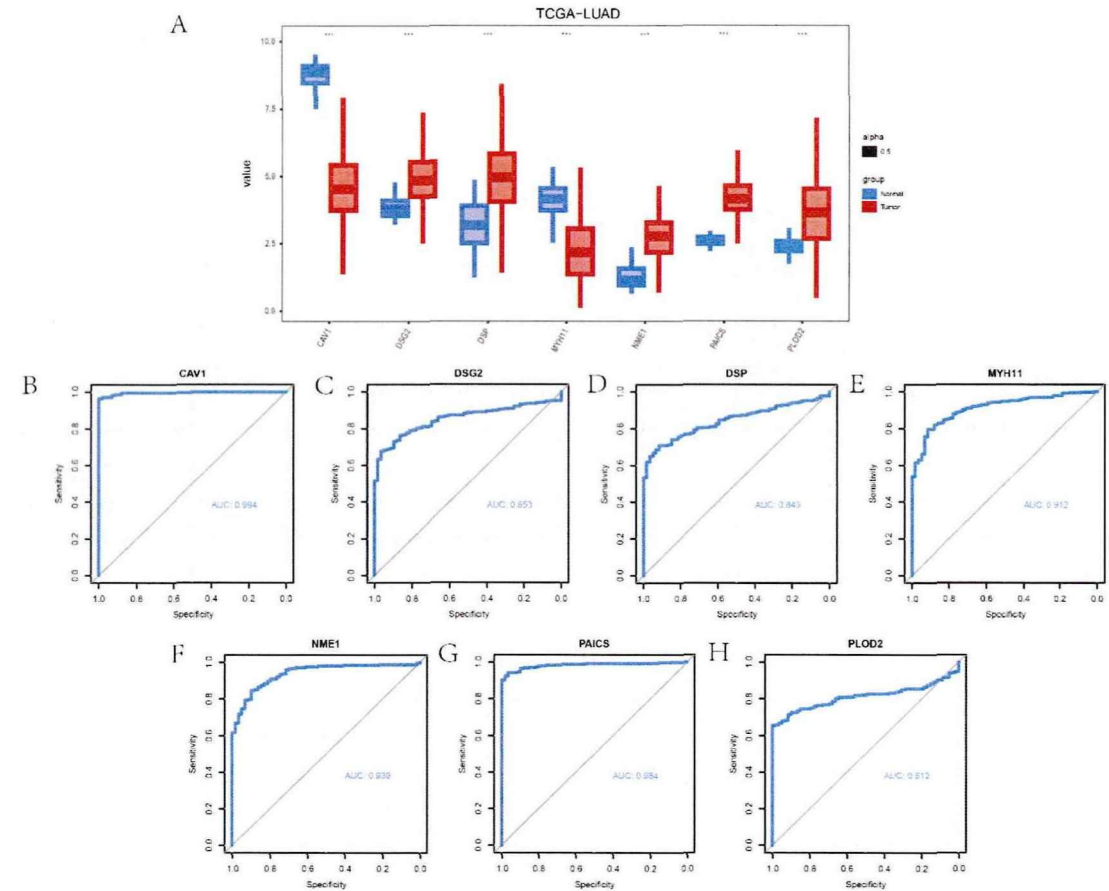


图 8. TCGA-LUAD 数据集中 MiAg DEGs 的表达差异分析

A. MiAg DEGs 的在 TCGA-LUAD 数据集中的表达差异分析分组比较图。B-S. MiAg DEGs 的 CAV1 (B), DSG2 (C), DSP (D), MYH11 (E), NME1 (F), PAICS (G), PLOD2 (H) 在 TCGA-LUAD 数据集中的 ROC 曲线结果展示。(*, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$)

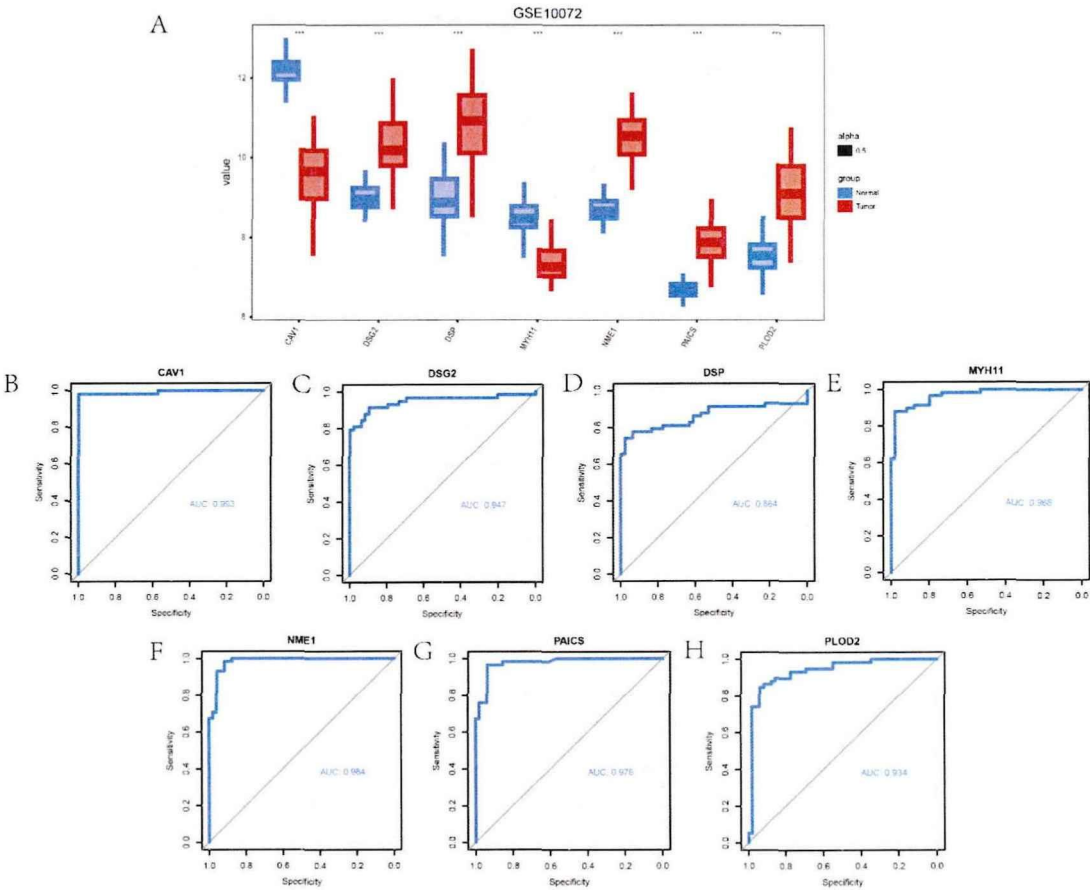


图 9. GSE10072 数据集中 MiAg DEGs 的表达差异分析

A. MiAg DEGs 的在 GSE10072 数据集中的表达差异分析分组比较图。B-S. MiAg DEGs 的 CAV1 (B), DSG2 (C), DSP (D), MYH11 (E), NME1 (F), PAICS (G), PLOD2 (H) 在 GSE10072 数据集中的 ROC 曲线结果展示。(*, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$)

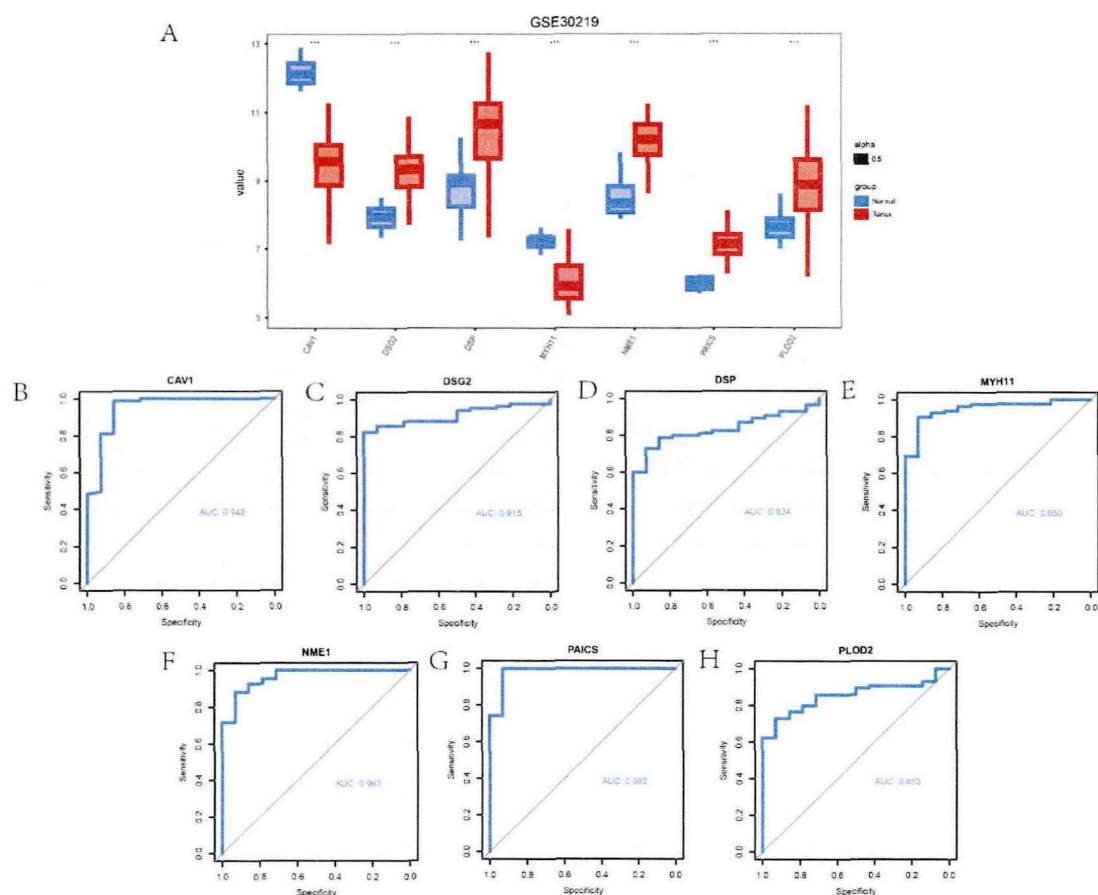


图 10. GSE30219 数据集中 MiAg DEGs 的表达差异分析

A. MiAg DEGs 的在 GSE30219 数据集中的表达差异分析分组比较图。B-S. MiAg DEGs 的 CAV1 (B), DSG2 (C), DSP (D), MYH11 (E), NME1 (F), PAICS (G), PLOD2 (H) 在 GSE30219 数据集中的 ROC 曲线结果展示。(*, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$)

六, MiAg DEGs 功能富集分析

我们针对 7 个 MiAg DEGs (包括 CAV1、DSG2、DSP、MYH11、NME1、PAICS、PLOD2) 之间与 LUAD 相关的生物学过程、分子功能、细胞组分和生物学通路进行深入分析。通过进行 GO (基因本体) 基因功能富集分析 (见表 3), 我们设定了筛选标准: $P_{adj} < 0.05$ 以及 FDR 值 ($q.value$) < 0.05 , 以确保分析结果具有统计学意义。结果显示 7 个 MiAg DEGs 在 LUAD 中主要富集在 regulation of ventricular cardiac muscle cell action potential, regulation of cardiac muscle cell action potential, regulation of cardiac muscle cell contraction, ventricular cardiac muscle cell action potential, regulation of heart rate by cardiac conduction 等生物学过程中, 以及 desmosome,

cornified envelope, intercalated disc, cell-cell contact zone, fascia adherens 等细胞组分中，同时还富集在 protein binding involved in heterotypic cell-cell adhesion, cell-cell adhesion mediator activity, cell adhesion mediator activity, nitric-oxide synthase binding, ribosomal small subunit binding 等分子功能中。我们通过气泡图对 GO 功能富集分析的结果进行了展示（图 11A）。另外，我们还将 GO 基因功能富集分析结果中的生物学过程通路、细胞组分通路和分子功能通路以网络图的形式展示出来（图 11B-D）。

我们对 7 个 MiAg DEGs 的 GO 功能进行了基于联合 logFC 值的富集分析。此分析不仅基于 GO 功能富集的标准方法，还额外引入了来自 TCGA-ESCC 数据集的 7 个 MiAg DEGs 的 logFC 值，以此计算每个基因的 z-score。这些联合 logFC 的 GO 功能富集分析结果通过气泡图形式展示，如图 11E 所示。

表 3. 线粒体自噬和衰老相关差异基因的功能富集分析

Ontology	ID	Description	GeneRatio	BgRatio	p value	p.adjust
生物学过程	GO:0098911	心室心肌细胞动作电位的调节	3/7	12/18800	6.94e-09	3.42e-06
生物学过程	GO:0098901	调节心肌细胞的动作电位	3/7	28/18800	1.03e-07	2.53e-05
生物学过程	GO:0086004	调节心肌细胞收缩	3/7	34/18800	1.88e-07	2.53e-05
生物学过程	GO:0086005	心室心肌细胞动作电位	3/7	35/18800	2.06e-07	2.53e-05
生物学过程	GO:0086091	通过心脏传导调节心率	3/7	39/18800	2.87e-07	2.55e-05
细胞组分	GO:0030057	桥粒	2/7	25/19594	3.27e-05	0.0015
细胞组分	GO:0001533	角质化包膜	2/7	45/19594	0.0001	0.0019
细胞组分	GO:0014704	闰盘	2/7	49/19594	0.0001	0.0019

Ontology	ID	Description	GeneRatio	BgRatio	p value	p.adjust
细胞组分	GO:0044291	细胞间接触区	2/7	71/19594	0.0003	0.0030
细胞组分	GO:0005916	粘附肌膜	1/7	10/19594	0.0036	0.0321
分子功能	GO:0086080	参与异型细胞-细胞粘附的蛋白质结合	2/7	12/18410	8.16e-06	0.0005
分子功能	GO:0098632	细胞-细胞粘附介质活性	2/7	54/18410	0.0002	0.0049
分子功能	GO:0098631	细胞粘附介质活性	2/7	64/18410	0.0002	0.0049
分子功能	GO:0050998	与一氧化氮合成酶结合	1/7	13/18410	0.0049	0.0565
分子功能	GO:0043024	结合核糖体小亚基	1/7	17/18410	0.0064	0.0565

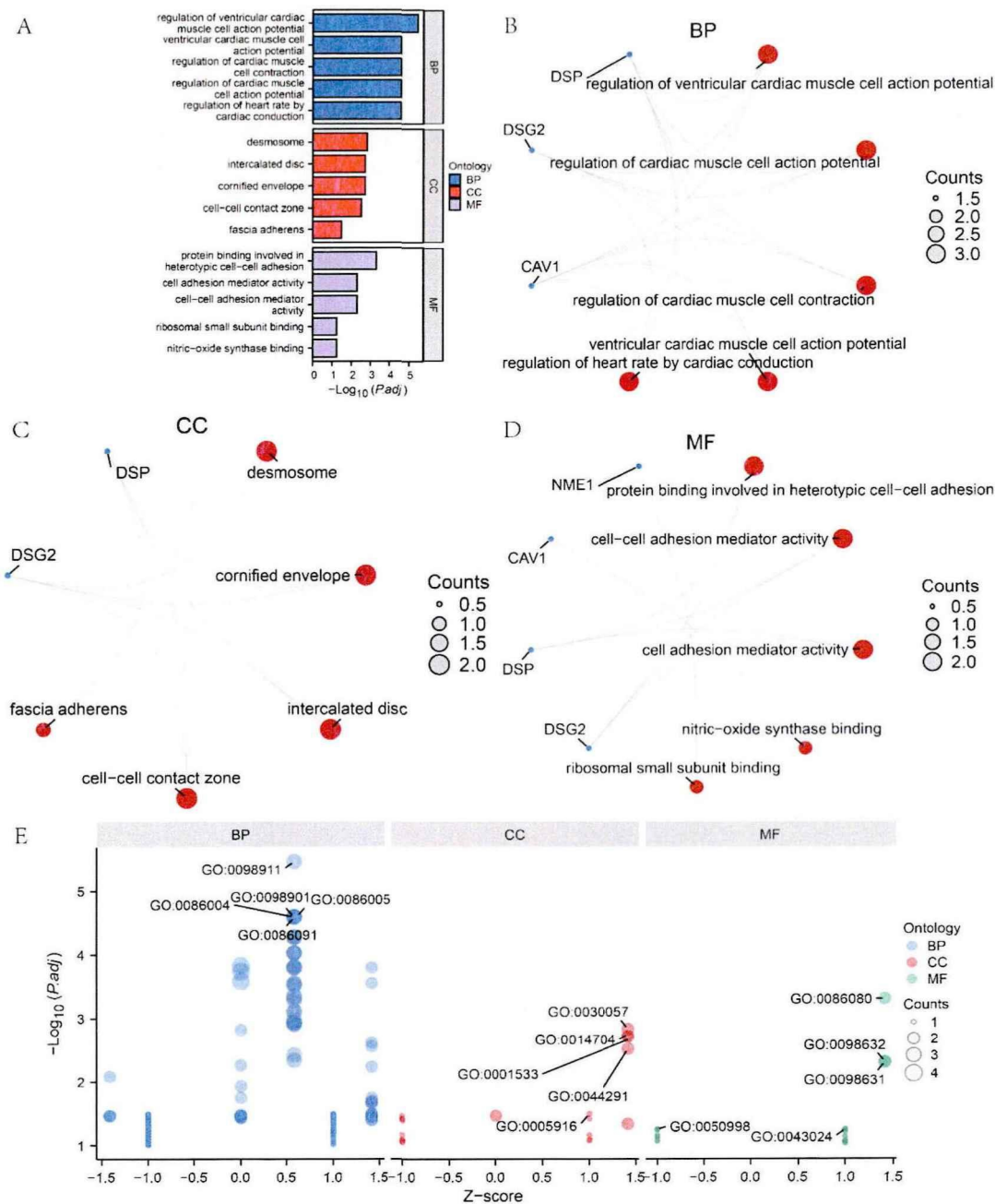


图 11. MiAg DEGs 功能富集分析

A. MiAg DEGs 的 GO 功能富集分析结果柱状图展示。B-D. MiAg DEGs 的 GO 功能富集分析结果中生物学过程通路 (B)，细胞组分通路 (C)，分子功能通路 (D) 的环形网络图展示。E. MiAg DEGs 联合 logFC 的 GO/KEGG 富集分析结果气泡图展示。

七, LUAD 数据集的基因集富集分析

为了探究基因表达水平对 LUAD 发病机制的影响,我们采用了基因集富集分析 (GSEA) 来评估 TCGA-LUAD 数据集中的所有基因表达与它们所涉及的生物学过程、细胞组分及分子功能之间的关系。本研究设置了 P 调整值小于 0.05 和 FDR 值小于 0.25 为显著富集的阈值条件。结果显示,TCGA-LUAD 数据集中 MiAg Scores 高低评分分组的基因显著富集在 FCERI 蛋白所介导的 MAPK 通路激活 (图 12B), G2 期-M 期 DNA 损伤检查点 (图 12C), FCERI 蛋白所介导的 NF- κ B 通路激活 (图 12D), The β -catenin/TCF 复合体的形成 (图 12E), Pre-notch 的表达与处理 (图 12F) 等通路中 (图 12B-F, 表 4)。

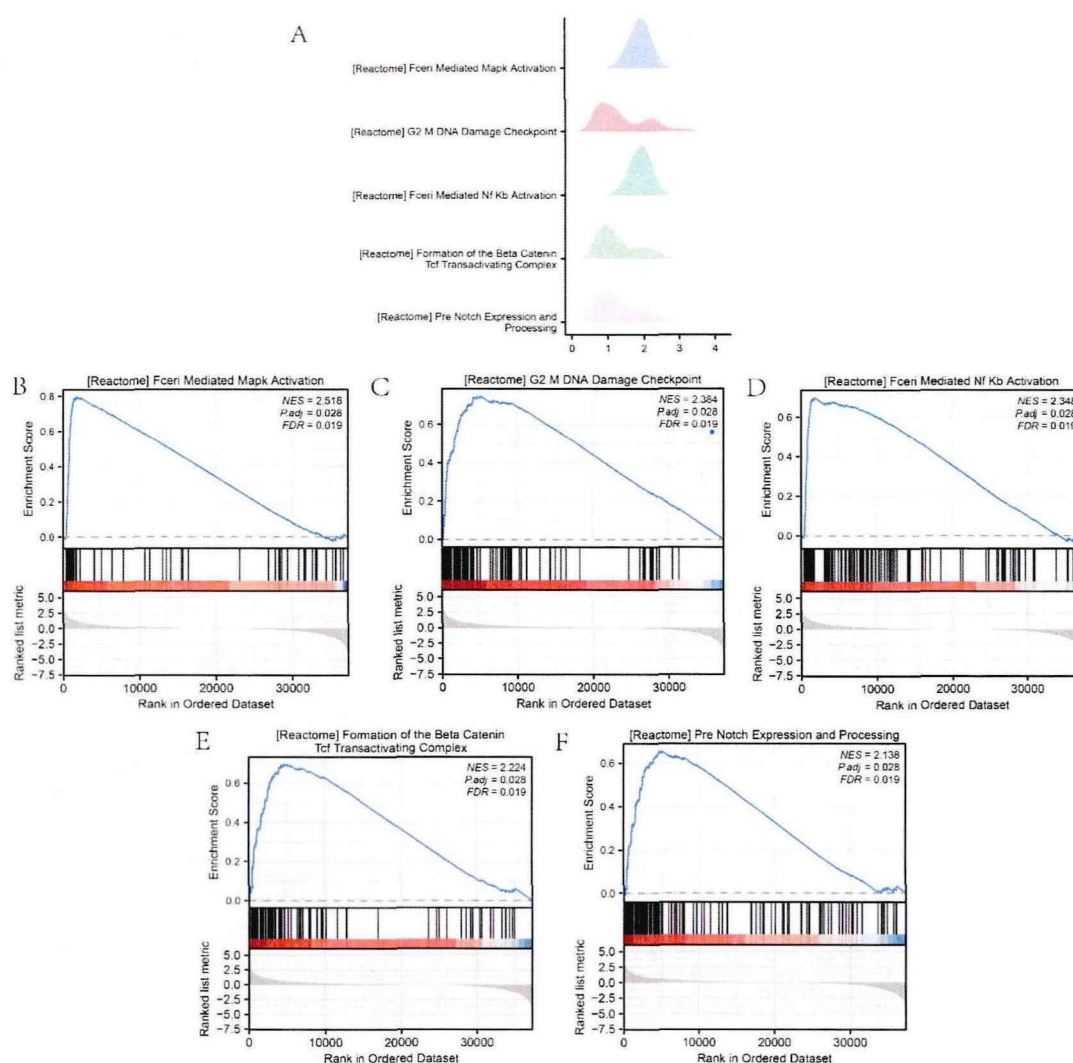


图 12. LUAD 数据集的基因集富集分析

A, TCGA- LUAD 数据集的 GSEA 富集分析主要 5 项生物学特征。B-E. TCGA-LUAD 数据集中的差异表达基因显著富集在 FCERI 蛋白所介导的 MAPK 通路激活 (B), G2 期-M 期 DNA 损伤检查点 (C), FCERI 蛋白所介导的 NF κ B 通路激活 (D),

The β -catenin/TCF 复合体的形成 (E), Pre-norch 的表达与处理 (F) 通路。

表 4. TCGA-LUAD 数据集中的基因集富集分析

Description	setSize	enrichmentScore	NES	p.adjust	qvalue
FCERI 蛋白所介导的 MAPK 通路激活	87	0.795790298	2.517874907	0.027730132	0.019016072
G2 期-M 期 DNA 损伤检查点	93	0.749162616	2.383853402	0.027730132	0.019016072
FCERI 蛋白所介导的 NF κ B 通路激活	136	0.699947529	2.347541845	0.027730132	0.019016072
The β-catenin/TCF 复合体的形成	91	0.698582462	2.22447468	0.027730132	0.019016072
Pre-norch 的表达与处理	118	0.656529438	2.137815726	0.027730132	0.019016072
氧化应激诱导的衰老	123	0.637050316	2.096642527	0.027730132	0.019016072
结肠癌中的代谢重编程	42	0.647220895	1.783359701	0.032411453	0.022226311
TP53 介导的反应组转录调节	360	0.467896572	1.732628717	0.032149094	0.022046397
G1 期-S 期 DNA 损伤检查点	68	0.564444252	1.705167645	0.032411453	0.022226311
NOTCH 信号通路	245	0.477834874	1.693454307	0.028653393	0.019649203

八, MiAg DEGs 预后模型的构建以及高低风险分组间的差异分析

为了评估 7 个 MiAg DEGs (CAV1、DSG2、DSP、MYH11、NME1、PAICS、PLOD2) 在 TCGA-LUAD 数据集中的预后意义, 我们采用了基于 LASSO 回归的方法来构建这些基因的预后模型 (见图 13A)。此外, 对 LASSO 回归的结果进行了可视化处理, 展示了基因的 LASSO 选择路径 (图 13B)。从这一分析中, 我们确定了模型中重要的 5 个基因: CAV1、DSG2、DSP、MYH11 和 PLOD2。

$$\text{RiskScore} = \text{CAV1} * 0.098 + \text{DSG2} * 0.188 + \text{DSP} * -0.030 + \text{MYH11} * 0.235 \\ + \text{PLOD2} * 0.028$$

接着, 我们通过风险因子图来展示 MiAg DEGs 构建的预后模型样本的分类情况 (见图 13C)。该风险因子图包括三个部分: 第一部分为风险分组, 根据从 TCGA-LUAD 数据集得到的样本的风险得分 (Risk Score), 使用中位数进行分组; 第二部分为生存结局的展示, 使用点图显示样本的生存时间和状态; 第三部分为热图, 展示了在预后模型样本中 MiAg DEGs 预后模型中所选基因的表达情况。

为了评估在 TCGA-LUAD 数据集中, MiAg DEGs 预后模型 RiskScore 高低风险组的基因表达差异, 我们应用了 limma 软件包, 对数据集进行了分析, 从而识别出在高低 RiskScore 风险组之间表达有显著差异的基因 (Differentially expressed genes, DEGs)。结果如下: TCGA-LUAD 数据集共得到了 56683 个差异表达基因, 其中满足 $|\log\text{FC}| > 1$ 且 $P. \text{adj} < 0.05$ 阈值的基因有 610 个, 在这一阈值下, 340 个在高 RiskScore 分组中高表达 (低 RiskScore 分组中低表达, $\log\text{FC}$ 为正, 上调基因), 270 个在高 RiskScore 分组中低表达 (低 RiskScore 分组中高表达, $\log\text{FC}$ 为负)。我们对 TCGA-LUAD 数据集高低 RiskScore 分组的差异分析结果绘制了火山图 (图 13F)。

我们使用 R 包 pheatmap 绘制热图展示 TCGA-LUAD 数据集 MiAg DEGs 预后模型高低风险组中的 5 个预后 MiAg RGs 的差异分析结果 (图 13G)。

我们对 MiAg DEGs 预后模型的 RiskScore 绘制预后生存 KM 曲线, 结果显示 MiAg DEGs 预后模型的 RiskScore 在 TCGA-LUAD 数据集 LUAD 患者生存结局即总生存 (Overall Survival, OS) 的预测具有较高的准确性 ($P < 0.001$, 图 13F)。

另外, 我们还绘制了时间依赖 ROC 曲线分析 MiAg DEGs 预后模型的 RiskScore 与 TCGA-LUAD 数据集 LUAD 患者生存信息 (PFI.event) 之间的相关性。结果显示 MiAg DEGs 预后模型的 RiskScore 在 TCGA-LUAD 数据集 LUAD 患者生存结局 (OS) 的预测具有较低的准确性 ($\text{AUC1}=0.641, \text{AUC3}=0.651, \text{AUC5}=0.600$, 图 13G)。

为了分析 MiAg DEGs 预后模型在 GSE3141 数据集中的预测准确性, 我们将 LASSO 模型系数与 GSE3141 数据集各样本的基因表达量相乘并累加, 获得 GSE3141 数据集各样本的 RiskScore。

我们对 MiAg DEGs 预后模型的 RiskScore 绘制预后生存 KM 曲线，结果显示 MiAg DEGs 预后模型的 RiskScore 在 GSE3141 数据集 LUAD 患者生存结局（OS）的预测具有较高的准确性（ $P<0.05$ ，图 13H）。

另外，我们还绘制了时间依赖 ROC 曲线分析 MiAg DEGs 预后模型的 RiskScore 与 GSE3141 数据集 LUAD 患者生存信息之间的相关性。结果显示 MiAg DEGs 预后模型的 RiskScore 在 GSE3141 数据集 LUAD 患者 1 年、3 年和 5 年生存结局（OS）的预测具有一定的准确性（ $AUC1=0.706$ ， $AUC3=0.752$ ， $AUC5=0.692$ ，图 13I）。

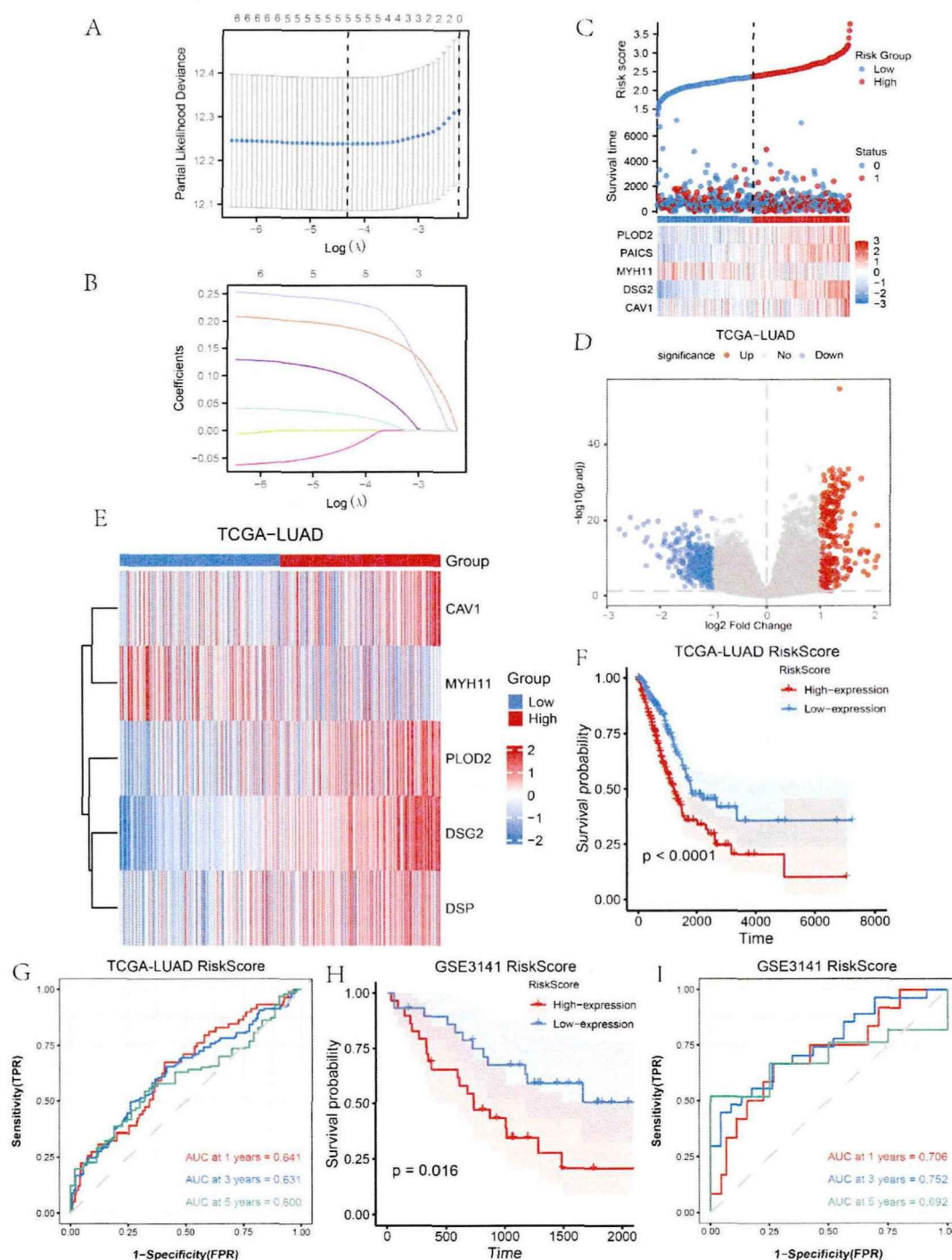


图 13. MiAg DEGs 预后模型的构建与高低风险评分分组间的差异分析

A. TCGA- LUAD 数据集 MiAg DEGs 的 LASSO 回归预后模型图。B-C. LASSO 回归预后模型的变量轨迹图(B), 风险因子图(C)。D. TCGA-LUAD 数据集 MiAg DEGs 预后模型高低 RiskScore 分组间的差异分析结果火山图。E. TCGA-LUAD 数据集 MiAg DEGs 预后模型高低风险组间的差异分析结果热图。F. TCGA-LUAD 数据集

MiAg DEGs 预后模型 RiskScore 高低风险组的预后分析 KM 曲线。G. MiAg DEGs 预后模型 RiskScore 高低风险组基于 TCGA-LUAD 数据集肺腺癌 (LUAD) 患者生存结局的时间依赖 ROC 曲线。H. GSE3141 数据集 MiAg DEGs 预后模型 RiskScore 高低风险组的预后分析 KM 曲线。I. MiAg DEGs 预后模型 RiskScore 高低风险组基于 GSE3141 数据集 LUAD 患者生存结局的时间依赖 ROC 曲线。(*, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$)

九、高低风险组间差异表达基因的基因集变异分析

为了探究 hallmark 基因集在 TCGA-LUAD 数据集高低风险分组间的差异性, 随后我们对 TCGA-LUAD 数据集高低风险分组间差异表达基因进行了基因集变异分析 (Gene Set Variation Analysis, GSVA) (图 14, 表 5)。TCGA-LUAD 数据集高低风险分组间的 GSVA 结果显示, G2 期-M 期检查点, MYC targets V1 (细胞增殖相关), mTORC1 信号通路, E2F-targets 通路, 有丝分裂纺锤体等共 26 个 hallmark 基因集在 TCGA-LUAD 数据集高低风险组间显示出显著差异 ($P_{\text{adj.}} < 0.05$), 其中雄激素应答, 凝血, DNA 修复等共 17 条通路在 TCGA-LUAD 数据集高风险组中的富集得分显著高于低风险组 ($P_{\text{adj.}} < 0.05$), 而胆汁酸代谢, Hedgehog 信号通路, 血红素代谢等其他 9 条通路在 TCGA-LUAD 数据集低风险组的富集得分显著高于高风险组 ($P_{\text{adj.}} < 0.05$)。我们根据 GSVA 得到的结果对 26 条 hallmark 通路在 TCGA-LUAD 数据集不同分组间的差异表达情况进行分析并使用 R 包 pheatmap 绘制热图展示具体差异分析结果(图 14A)。我们通过 Mann-Whitney U 检验分析 TCGA-LUAD 数据集不同分组间 26 条 hallmark 通路的分组差异程度并通过分组比较图进行结果展示(图 14B), 结果显示 26 条 hallmark 通路在 TCGA-LUAD 数据集不同分组间的差异表达具有显著统计学差异 ($P < 0.05$)。

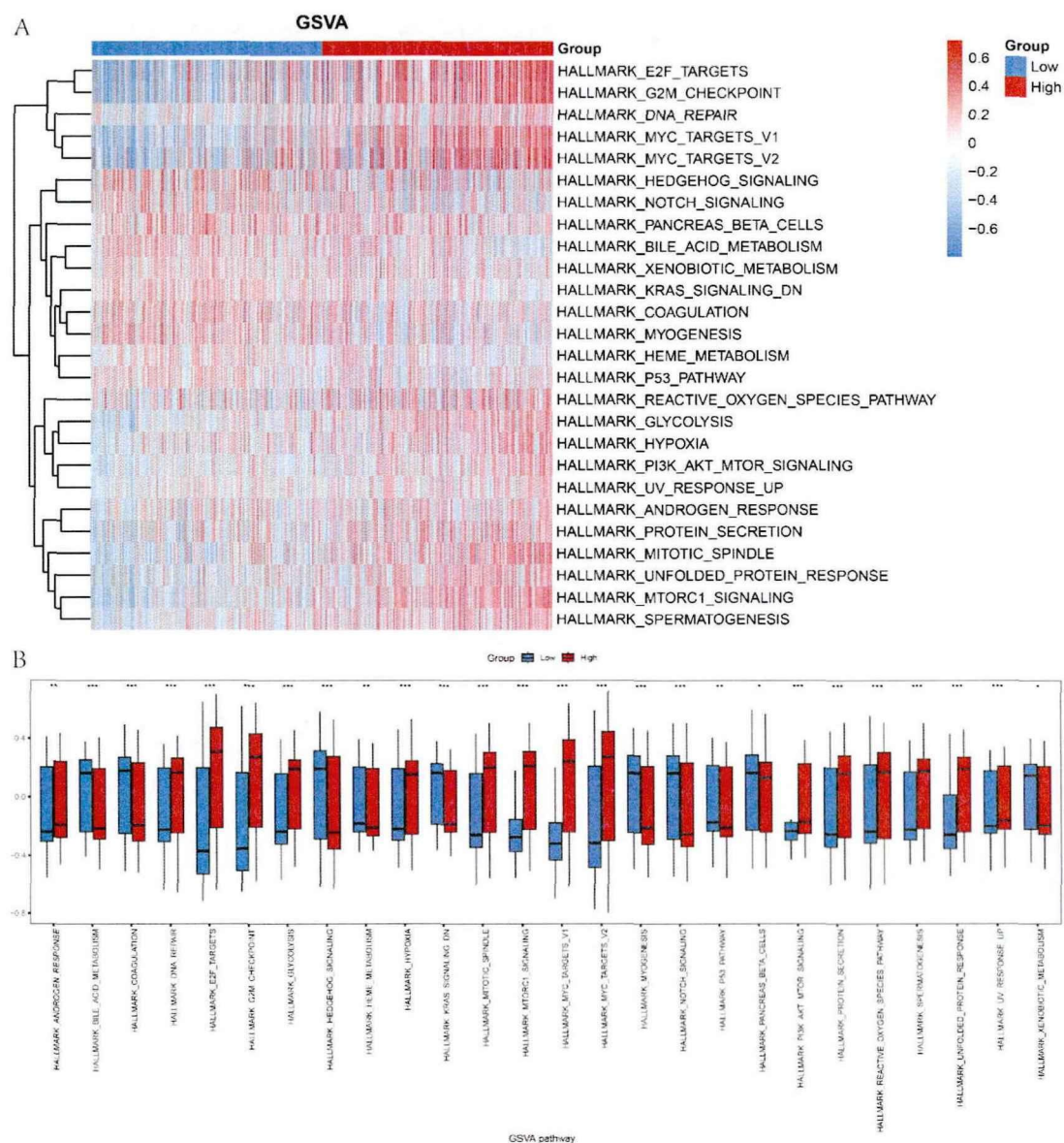


图 14. 高低风险组间 DEGs 的基因集变异分析

A. TCGA-LUAD 数据集 MiAg DEGs 预后模型高低风险组间 DEGs 的 GSVA 富集分析结果的热图;B. TCGA-LUAD 数据集 MiAg DEGs 预后模型高低风险组间的 GSVA 富集分析结果的分组比较图。（*, P < 0.05; **, P < 0.01; ***, P < 0.001）

表 5. MiAg DEGs 预后模型中高低评分分组间的基因组变异分析

Description	logFC	AveExpr	t	adj.P.Val
G2 期-M 期检查点	-0.373005414	-0.038347037	-12.22174294	7.27E-29
MYC targets V1	-0.339251292	-0.052047523	-12.11581284	1.01E-28
mTORC1 信号通路	-0.277280327	-0.049483713	-11.68582277	4.06E-27

E2F-targets 通路	-0.377802114	-0.032942337	-11.30320163	1.09E-25
有丝分裂纺锤体	-0.234010455	-0.051965342	-9.423953801	1.19E-18
MYC targets V2	-0.301275477	-0.041007769	-8.956982377	4.27E-17
GLYCOLYSIS	-0.190942406	-0.055941112	-8.710421032	2.53E-16
未折叠蛋白反应	-0.199586725	-0.067364522	-8.584379032	5.84E-16
精子生成	-0.15183207	-0.039303228	-6.863210786	1.01E-10
PI3K-AKT-MTOR 信号通路	-0.13168222	-0.086085896	-6.424165862	1.43E-09
KRAS 信号通路	0.117245966	-0.013312682	5.949485802	2.17E-08
蛋白分泌	-0.140660104	-0.064838007	-5.532970938	2.03E-07
肌肉生成	0.128482952	-0.036397891	5.248337734	8.44E-07
DNA 修复	-0.123740423	-0.058708012	-5.211710483	9.46E-07
胆汁酸代谢	0.114989964	-0.036807337	5.093390069	1.61E-06
缺氧	-0.101643691	-0.046178241	-4.308218689	6.09E-05
NOTCH 信号通路	0.108867274	-0.056773358	4.143311959	0.000116512
凝血	0.095527137	-0.017508103	3.822276203	0.000409159
ROS 通路	-0.097034727	-0.047674107	-3.603143502	0.000902232
UV-response-up	-0.069027027	-0.0584241	-3.422326367	0.001667428
Hedgehog 信号通路	0.098517204	-0.026335964	3.379090581	0.001854033
雄激素应答	-0.071030058	-0.067485189	-3.008895785	0.006231461
血红素代谢	0.055124647	-0.067813937	2.74396608	0.013626014
P53 通路	0.057092792	-0.053322997	2.597963864	0.020059655
外源物质代谢	0.052817176	-0.030101536	2.414717822	0.03214407
胰岛 β 细胞	0.054909713	0.033381435	2.304046158	0.041521419

十，线粒体自噬和衰老表型评分的构建与免疫治疗评估

我们基于 5 个 MiAg DEGs 在 TCGA-LUAD 数据集中表达情况，通过单样本基因集富集分析（single-sample gene-set enrichment analysis, ssGSEA）算法获取了 TCGA-LUAD 数据集中每个 LUAD 患者的线粒体自噬和衰老表型评分(MiAg Scores)以代表 LUAD 患者的 MiAg RGs 表达水平。然后将 TCGA-LUAD 数据集中 LUAD 患者样本的 MiAg Scores 根据中位数分成高，低评分两组。然后我们结合 TCGA-LUAD 数据集 LUAD 患者的预后信息对 MiAg Scores 在 TCGA 数据集不同预后结局中通过绘制预后 KM 曲线计算以确定 MiAg Scores 在 TCGA-LUAD 数据

集中的最佳分组（中位数分组），并根据最佳分组将 TCGA-LUAD 数据集 LUAD 患者样本的 MiAg Scores 分成高，低评分两组， P value < 0.05 被认为具有统计学差异（图 15A）。由图 15A 可知，MiAg Scores ($P=0.028$, 图 15A) 在 TCGA-LUAD 数据集中的表达水平对 LUAD 患者的预后 PFI 的预测结果存在一定的准确性 ($P<0.05$)。

我们对不同预后结局间的 MiAg Scores 的分组差异情况进行分析，我们使用 WilCoxon 秩和检验分析了 MiAg Scores 在 TCGA-LUAD 数据集中表达量高低在不同预后结局间的差异并通过分组比较图将表达差异分析结果展示出来（图 15B）。MiAg Scores 在 TCGA-LUAD 数据集不同预后结局间的表达具统计学显著差异 ($P < 0.01$)（图 15B）。

我们对 RiskScore 高低风险组间的 MiAg Scores 的分组差异情况进行分析，我们使用 WilCoxon 秩和检验分析了 MiAg Scores 在 TCGA-LUAD 数据集中表达量高低在高低风险组间的差异并通过分组比较图将表达差异分析结果展示出来（图 15C）。MiAg Scores 在 TCGA-LUAD 数据集高低风险组间的表达具统计学显著差异 ($P < 0.01$)（图 15C）。然后我们基于 Riskscore 与 MiAg Scores 进行相关性分析，结果显示：TCGA-LUAD 数据集 MiAg DEGs 预后模型高低风险组间 Riskscore 与 MiAg Scores ($R=0.606$, $P < 0.001$, 图 15D) 之间存在中等正相关性。

为进一步分析 MiAg Scores 对免疫治疗的预测情况，我们通过 TCIA 数据库下载 LUAD 相关的免疫原性评分 (Immunophenoscore, IPS)，我们使用 WilCoxon 秩和检验分析了 IPS 在 TCGA-LUAD 数据集 MiAg Scores 高低分组间的差异并通过分组比较图将表达差异分析结果展示出来（图 15E-H）。结果显示，IPS 在 TCGA-LUAD 数据集 MiAg Scores 高低分组之间的差异极具统计学显著意义 ($P < 0.001$, 图 15E)，IPS-PD1 在 TCGA-LUAD 数据集 MiAg Scores 高低分组之间的差异无统计学显著意义 ($P > 0.05$, 图 15F)，IPS-CTLA4 在 TCGA-LUAD 数据集 MiAg Scores 高低分组之间的差异具有显著的统计学显著意义 ($P < 0.001$, 图 15G)，IPS-PD1-CTLA4 在 TCGA-LUAD 数据集 MiAg Scores 高低分组之间的差异无统计学显著意义 ($P > 0.05$, 图 15H)。

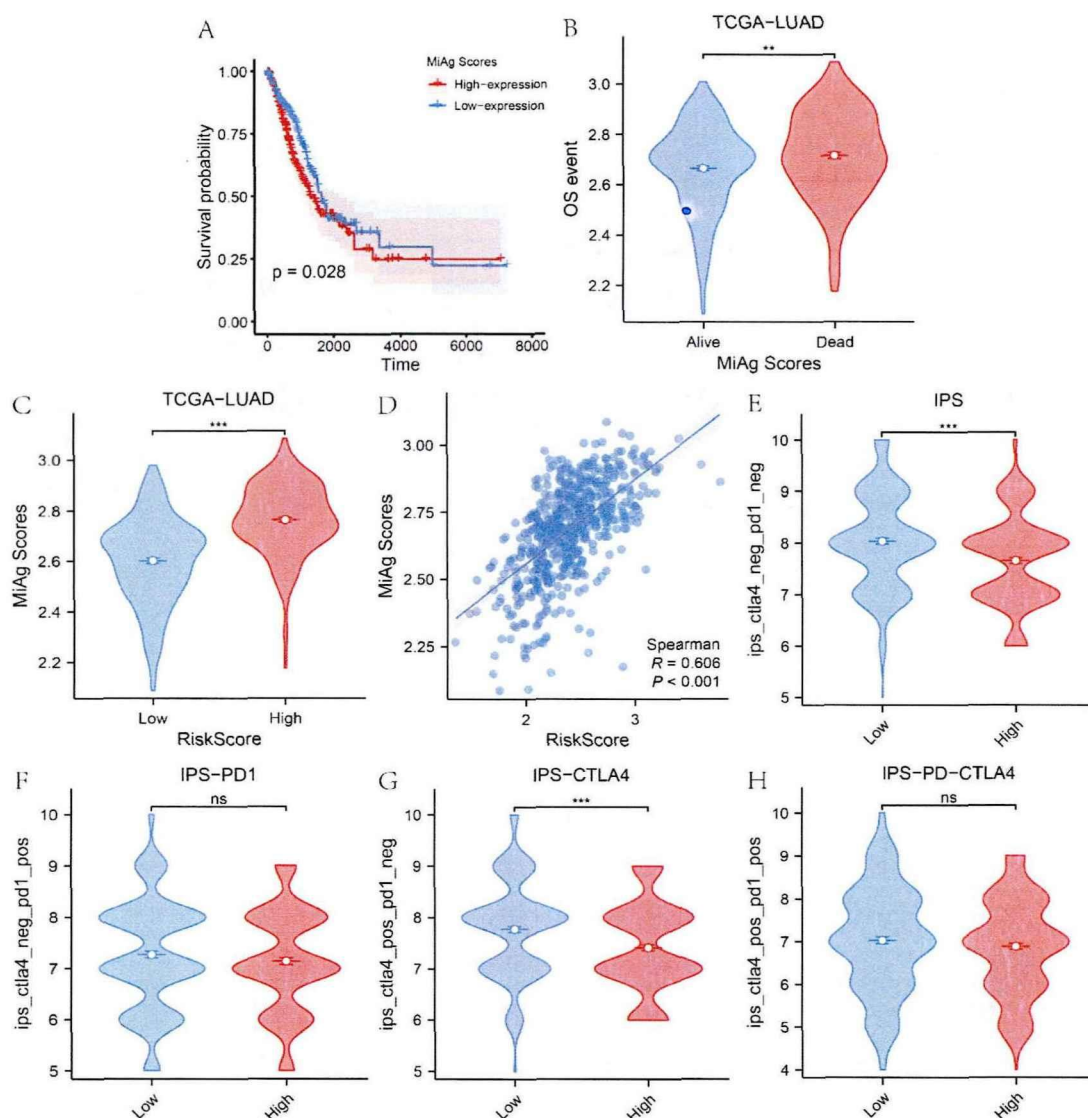


图 15. MiAg Scores 的构建与高低风险评分分组间的差异分析

A. MiAg Scores 在 TCGA-LUAD 数据集中的 KM 曲线结果展示。B. TCGA-LUAD 数据集不同预后结局间 MiAg Scores 的分组比较图。C. TCGA-LUAD 数据集 MiAg DEGs 预后模型 RiskScore 高低风险组间 MiAg Scores 的分组比较图。D. TCGA-LUAD 数据集 MiAg DEGs 预后模型 RiskScore 与 MiAg Scores 评分的相关性分析结果展示。E-G. TCGA-LUAD 数据集 MiAg Scores 高低组间 IPS (E), IPS-PD1 (F), IPS-CTLA4 (G), IPS-PD1-CTLA4 (H) 的分组比较图。(*, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$)

十一, MiAg Scores 高低评分分组之间的免疫特征差异分析

随后我们对 TCGA-LUAD 数据集 MiAg Scores 高低评分分组样本使用了 CIBERSORT 算法计算了 22 种免疫细胞在高低评分组中的浸润丰度, 然后通过 Mann-Whitney U 检验分析高低评分分组间 22 种免疫细胞的浸润丰度差异并通过分组比较图进行结果展示 (图 16A)。结果显示在 TCGA-LUAD 数据集 MiAg Scores 高低评分分组间中有 8 种免疫细胞浸润丰度在不同分组间具有统计学显著差异 ($P < 0.05$), 分别是 B cells naive, Macrophages M0, Macrophages M1, Neutrophils, NK cells resting, T cells CD4 memory activated, T cells follicular helper, T cells regulatory (Tregs)。

然后我们分别计算了 MiAg Scores 高低评分组样本中 8 种免疫细胞浸润丰度之间的相关性并进行结果展示 (图 16B-C), 结果显示在低评分组样本 (图 16B) 和高评分组样本 (图 16C) 中, 10 种免疫细胞中的多数免疫细胞相互之间的浸润丰度具有一定的负相关性。

同时我们分别计算了低评分组样本 (图 16D) 和高评分组样本 (图 16E) 组患者样本中 8 种免疫细胞浸润丰度与 7 个 MiAg DEGs (CAV1, DSG2, DSP, MYH11, NME1, PAICS, PLOD2) 表达量之间的相关性 (图 16D-E), 结果显示在低评分组样本中免疫细胞 Macrophages M1 与 MYH11; T cells CD4 memory activated 与 MYH11; T cells regulatory (Tregs) 与 CAV1 之间的负相关性; 免疫细胞 Neutrophils 与 IL6 之间的正相关性较为明显。而高评分组样本中免疫细胞 T cells regulatory (Tregs) 与 MYH11, CYCS 之间的负相关性; 免疫细胞 Macrophages M0 与 PLOD2; Macrophages M1 与 NME1, PAICS, PLOD2; T cells CD4 memory activated 与 NME1, PAICS, PLOD2 之间的正相关性较为明显。

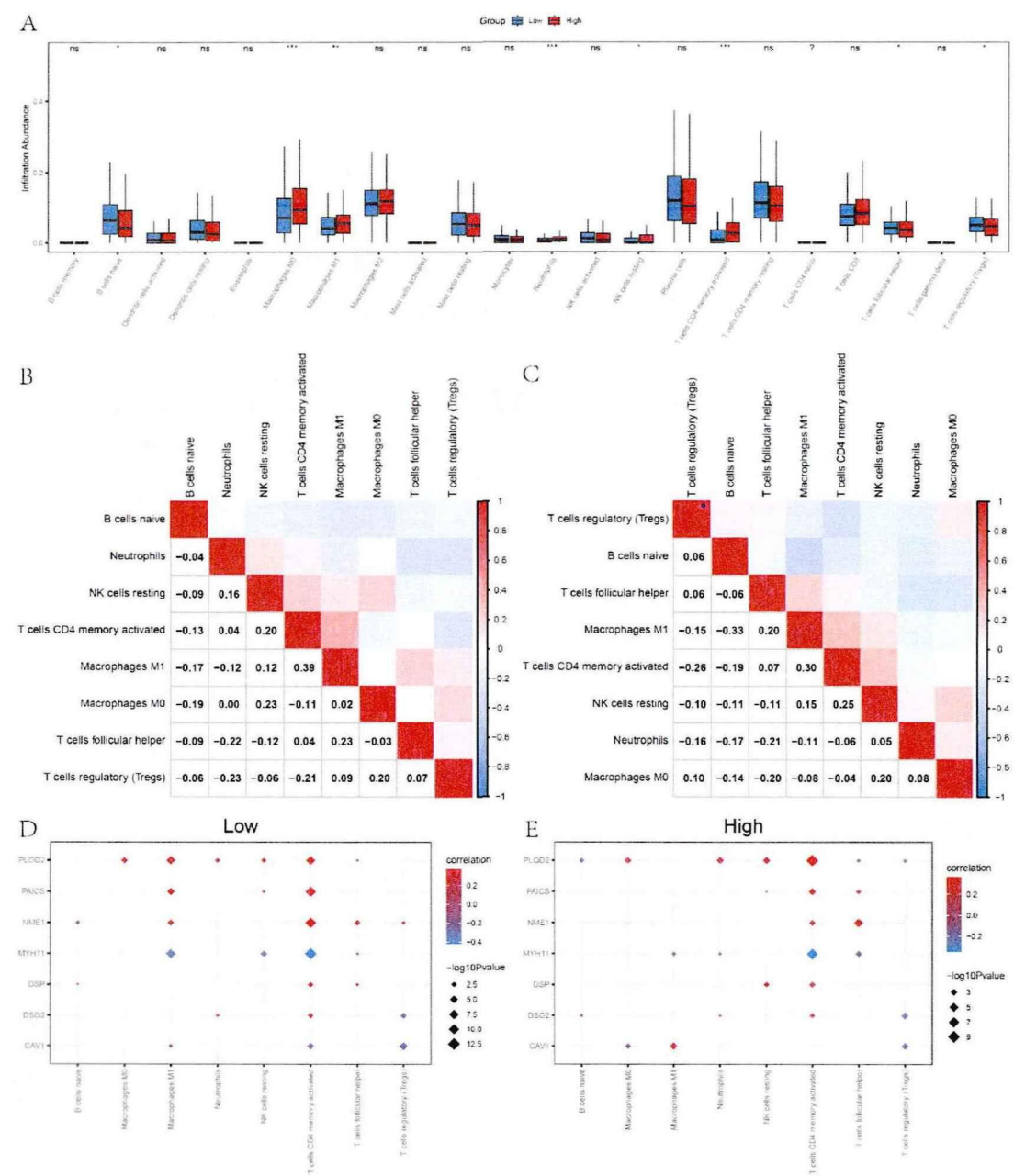


图 16. TCGA-LUAD 数据集 MiAg Scores 高低评分分组的免疫浸润分析

A.TCGA-LUAD 数据集 MiAg Scores 高低评分分组之间的 CIBERSORT 免疫浸润分析结果分组比较图展示。B-C. TCGA-LUAD 数据集 MiAg Scores 低评分分组 (B), 高分评分组 (C) 中的免疫细胞浸润丰度的相关性分析结果展示。D-E. TCGA-LUAD 数据集 MiAg Scores 低评分分组 (D), 高分评分组 (E) 中的免疫细胞浸润丰度的与 MiAg DEGs 的相关性热图。符号?表示表达量无法进行统计分析。(*, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$)

十二，TCGA-LUAD 数据集多因素 Cox 预后模型的构建

为进一步验证我们建立的 MiAg Scores 对 LUAD 患者预后的作用, 在 TCGA-LUAD 数据集内, 我们通过应用单变量和多变量 Cox 回归模型, 研究了 MiAg Scores 的风险评分和五种临床参数 (T 阶段, N 阶段, M 阶段, 年龄及性别) 对肺腺癌 (LUAD) 病患预后的影响。初步利用单变量分析探讨了 MiAg Scores 及临床参数, 随后将这些变量纳入到多变量分析中, 以建立一个综合的多变量 Cox 回归模型。研究发现, MiAg Scores 与患者的临床预后存在显著的相关性 (P 值小于 0.01, 见表 6)。此外, 我们将多变量 Cox 回归的结果以森林图形式呈现 (见图 17A)。

$$\text{RiskScore} = \text{Stage}_{T2} * 0.503 + \text{Stage}_{T3} * 1.255 + \text{Stage}_{T4} * 0.431 + \text{Stage}_{N1} * 0.740 + \text{Stage}_{N2} * 0.920 + \text{Stage}_{M1} * 0.745 + \text{MiAg Scores} * 1.521$$

随后, 我们进行了列线图分析以评估我们的模型预测患者生存率的能力, 并展示了相关图表 (图 17B)。分析结果表明, T 分期和 MiAg Scores 在多变量 Cox 回归模型中的影响力显著高于其他变量 (如图 17B 所示)。此外, 我们还对该模型进行了一年 (图 17C)、三年 (图 17D) 及五年 (图 17E) 的预后校正 (Calibration) 分析, 并绘制了相应的校准曲线图 (图 17C-E)。校准曲线图展示了模型预测的生存概率与实际生存概率之间的匹配程度, 通过不同颜色的线段和点标出不同时间点的预测情况。这些线段越接近灰色理想线, 表明模型在相应时间点的预测准确性越高。

我们接着应用决策曲线分析来评估我们建立的多因素 Cox 回归模型在 1 年 (图 17F)、3 年 (图 17G) 及 5 年 (图 17H) 的临床应用有效性, 并展示了这些结果 (图 17F-H)。在 DCA 图中, x 轴表示概率阈值 (Threshold Probability), 而 y 轴显示的是净收益。通过比较模型预测线与全阳性 (All positive) 和全阴性 (All negative) 的线, 可以判定模型的预测性能。线条的 x 值覆盖范围越广, 表明模型的预测效果越佳。分析结果表明, 我们的多因素 DEG 回归模型在预测临床结果方面表现为 5 年优于 3 年和 1 年。

表 6. TCGA-LUAD 数据集中 COX 回归分析

特征	病例数	单因素分析		多因素分析	
		Hazard ratio (95% CI)	P 值	Hazard ratio (95% CI)	P 值
T 分期	521		< 0.001		
T1	175	Reference		Reference	
T2	281	1.523 (1.070 - 2.169)	0.020	1.654 (1.050 - 2.605)	0.030
T3	47	2.971 (1.766 - 4.998)	< 0.001	3.509 (1.860 - 6.619)	< 0.001
T4	18	3.082 (1.587 - 5.986)	< 0.001	1.539 (0.711 - 3.331)	0.274
N 分期	509		< 0.001		
N0	343	Reference		Reference	
N1	95	2.325 (1.655 - 3.267)	< 0.001	2.096 (1.422 - 3.090)	< 0.001
N2	71	2.911 (1.994 - 4.248)	< 0.001	2.509 (1.619 - 3.890)	< 0.001
M 分期	375		0.010		
M0	351	Reference		Reference	
M1	24	2.179 (1.273 - 3.729)	0.004	2.106 (1.150 - 3.857)	0.016
性别	524		0.620		
女	282	Reference			
男	242	1.076 (0.807 - 1.434)	0.620		
年龄	524		0.894		
<=60	167	Reference			
>60	357	1.021 (0.750 - 1.391)	0.894		
MiAg Scores	526	4.818 (1.978 - 11.736)	< 0.001	4.576 (1.453 - 14.412)	0.009

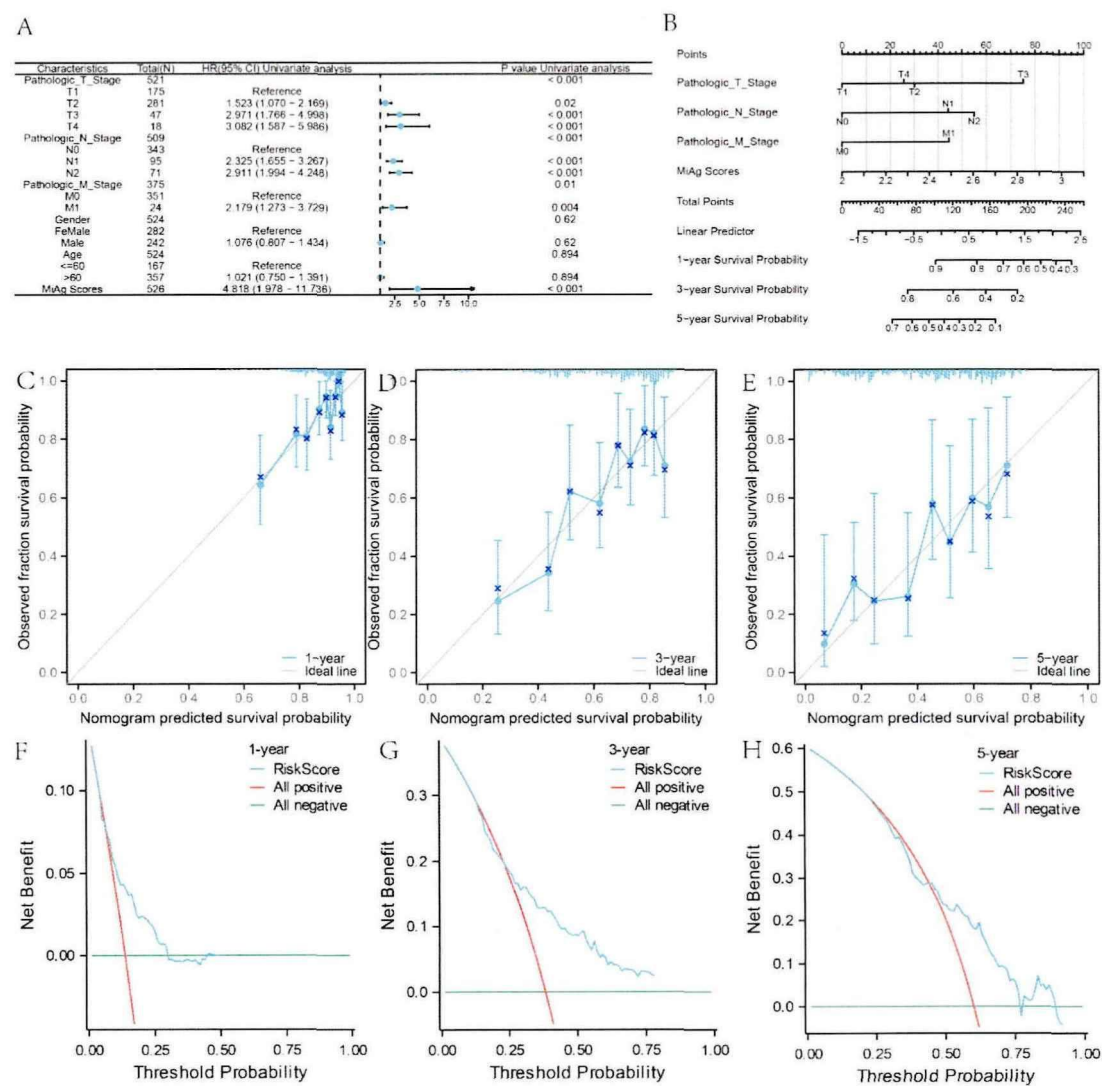


图 17. 高低评分组 MiAg Scores 的临床预后模型构建

A-B. TCGA-LUAD 数据集中 MiAg Scores 和临床变量的单因素及多因素 Cox 回归分析结果展示为森林图 (A) 和列线图 (B)。C-E. 根据多因素 Cox 回归模型, 进行的 nomogram 分析展示了 1 年 (C)、3 年 (D) 和 5 年 (E) 的校准曲线图。F-H. 这些模型在 1 年 (F)、3 年 (G)、5 年 (H) 的决策曲线分析 (DCA) 图表中, 用以显示预测的有效性和准确性。(*, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$)

十三, 本中心 LUAD 患者中构建预后模型的 MiAg DEGs 表达相关情况

接下来, 我们选取了就诊于中国医学科学院肿瘤医院胸外科并接受手术治疗的患者 22 例, 利用 Real-Time qPCR 技术, 对 LUAD 组织及其癌旁组织中 5 个 MiAg DEGs (PLOC2、DSG2、DSP、CAV1 及 MYH11) 的表达水平进行了定量分析。通

过提取 LUAD 组织和癌旁组织的总 RNA，采用 cDNA 合成和 Real-Time qPCR 扩增，我们成功测定了 5 个目标基因在 LUAD 组织及癌旁组织中的相对表达量。结果显示，CAV1（图 18A）以及 MYH11（图 18B）在癌旁组织中的表达量较 LUAD 组织中高，但两组未见明显统计学差异。相较于癌旁组织，LUAD 组织中 DSG2（图 18C）的表达水平显著上调（ $P<0.001$ ）；DSP（图 18D）在 LUAD 组织中的表达较癌旁组织低，但两组无统计学差异；PLOD2（图 18E）在 LUAD 组织中的表达较癌旁组织高，但两组无统计学差异。

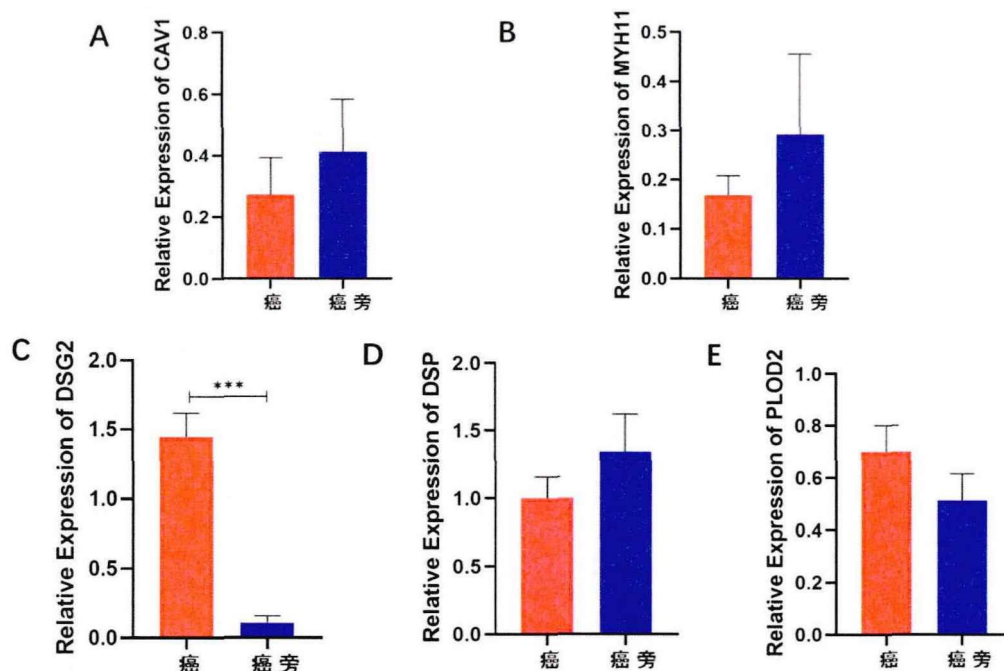


图 18. 本中心 LUAD 患者中 5 个 MiAg DEGs 的表达情况

A, RT-PCR 检测 CAV1 在 LUAD 组织和癌旁组织中的表达; B, RT-PCR 检测 MYH11 在 LUAD 组织和癌旁组织中的表达; C, RT-PCR 检测 DSG2 在 LUAD 组织和癌旁组织中的表达; D, RT-PCR 检测 DSP 在 LUAD 组织和癌旁组织中的表达; E, RT-PCR 检测 PLOD2 在 LUAD 组织和癌旁组织中的表达。(***, $p<0.001$)

十四, MiAg DEGs 在 LUAD 各类细胞中的表达和分布情况

为了更好地验证 MiAg DEGs 在 LUAD 各类细胞中的表达和分布情况，我们一共从公共数据库中选取了三个 LUAD scRNA-seq 数据集 (EMTAB6149, GSE127465, GSE143423) 进行分析。在分析中，我们一共将 LUAD 的细胞类型分为了四大类，分别为：恶性肿瘤细胞 (Malignant cells)、免疫细胞 (Immune cells)、基质细胞 (Stromal cells) 以及其他 (Others)。

在数据集 EMTAB6149 中,主要表达在恶性肿瘤细胞上的 MiAg DEGs 为 DSG2、DSP 以及 PLOD2, CAV1 主要表达在基质细胞和其他类型细胞中,在恶性肿瘤中表达量较低。而 MYH11 在四种类型的细胞中均未见明显的表达和分布(图 19A)。我们根据细胞类型构建 UMAP 图,将来自同一数据库样本中的 LUAD 组织单细胞数据进行细胞注释得到结果,并在图中用不同的颜色表示各种细胞分类(图 19B)。接下来我们将 MiAg DEGs 中的 5 个基因在各个细胞上的表达情况用颜色在 UMAP 图中标示出来,通过对照可以很直观地感受到 DSG2(图 19C)、DSP(图 19D)以及 PLOD2(图 19E)主要分布在恶性肿瘤细胞中,CAV1(图 19F)主要分布在基质细胞和其他类型的细胞中,而 MYH11(图 19G)只在少量基质细胞中分布。

在数据集 GSE127465 中,LUAD 细胞类型主要分为三大类型:恶性肿瘤细胞、免疫细胞和基质细胞。其中主要表达在恶性肿瘤细胞上的 MiAg DEGs 为 DSG2 和 DSP,CAV1 主要表达在基质细胞中,在恶性肿瘤细胞和免疫细胞中表达量较低。而 PLOD2 和 MYH11 仅在基质细胞中有少量的分布和表达,在其他类型的细胞中均未见明显的表达和分布(图 20A)。我们根据细胞类型构建 UMAP 图,将来自同一数据库样本中的 LUAD 组织单细胞数据进行细胞注释得到结果,并在图中用不同的颜色表示各种细胞分类(图 20B)。接下来我们将 MiAg DEGs 中的 5 个基因在各个细胞上的表达情况用颜色在 UMAP 图中标示出来,通过对照可以很直观地感受到 DSG2(图 20C)、DSP(图 20D)主要分布在恶性肿瘤细胞中,CAV1(图 20F)主要分布在基质细胞中,而 PLOD2(图 20E)和 MYH11(图 20G)只在少量基质细胞中分布。

在数据集 GSE143423 中,主要表达在恶性肿瘤细胞上的 MiAg DEGs 为 DSG2 和 DSP。CAV1 主要表达在基质细胞和其他类型的细胞中,在恶性肿瘤细胞表达量较低。而 PLOD2 仅在基质细胞和其他类型细胞中有少量的分布和表达,在恶性肿瘤细胞和淋巴细胞中均未见明显的表达和分布。MYH11 在四种类型的细胞中均未见明显的表达和分布(图 21A)。我们根据细胞类型构建 UMAP 图,将来自同一数据库样本中的 LUAD 组织单细胞数据进行细胞注释得到结果,并在图中用不同的颜色表示各种细胞分类(图 21B)。接下来我们将 MiAg DEGs 中的 5 个基因在各个细胞上的表达情况用颜色在 UMAP 图中标示出来,通过对照可以很直观地感受到 DSG2(图 21C)、DSP(图 21D)主要分布在恶性肿瘤细胞中,PLOD2(图 21E)和 CAV1(图 21F)在恶性肿瘤细胞中有少量的分布,而 MYH11(图 21G)在所有类型的细胞中未见明显的表达和分布。

结合以上的结果可以看出,在 LUAD 中,DSG2 和 DSP 主要分布和表达于恶性肿瘤细胞中。PLOD2 和 CAV1 虽然少量表达于恶性肿瘤细胞中,但主要分布于基质

细胞和其他类型的细胞中。而 MYH11 则在 LUAD 主要的细胞类型中未见规律的表达和分布。

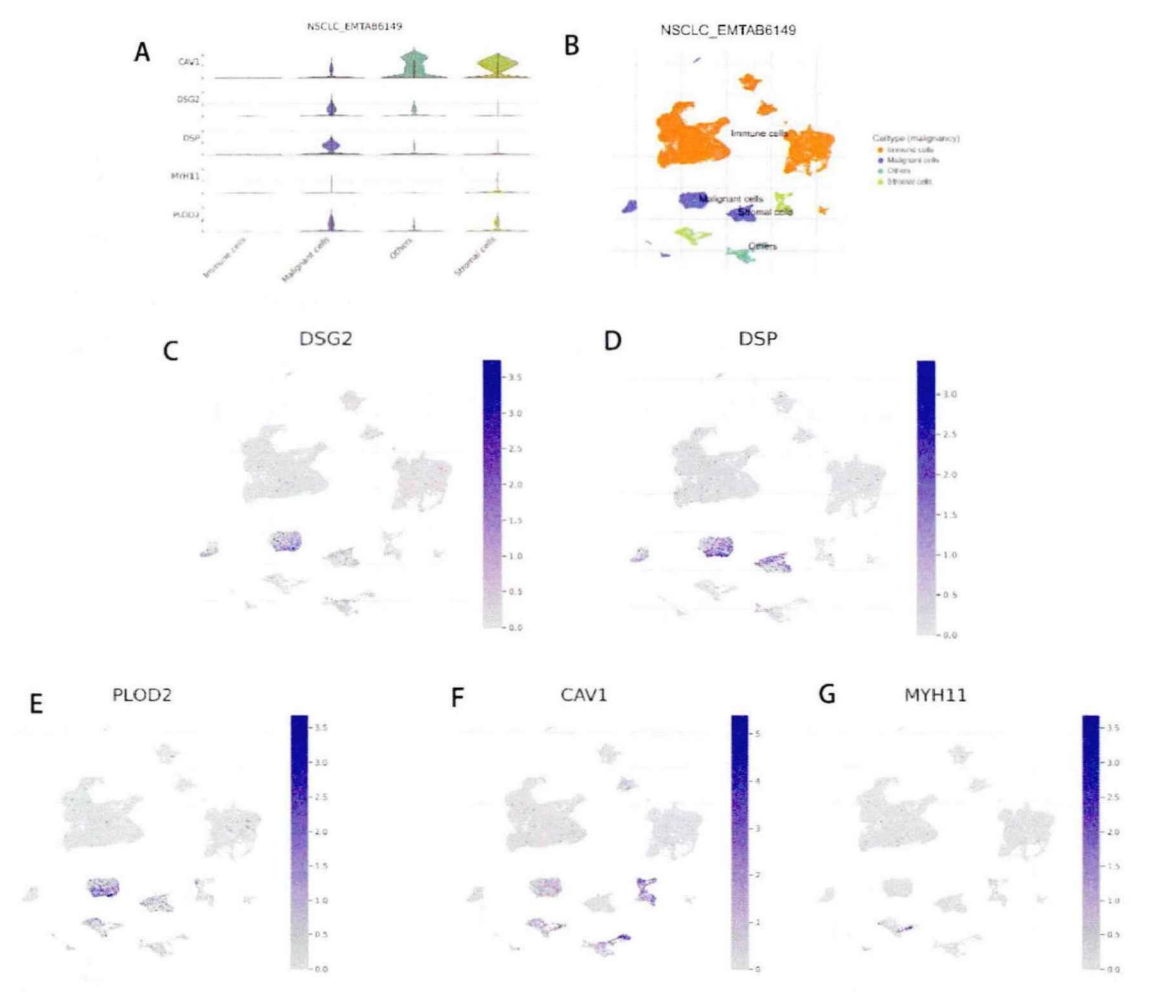


图 19. EMTAB6149 数据集中 MiAg DEGs 表达和分布的情况

A, MiAg DEGs 在 LUAD 四种主要类型细胞上的表达情况; B, LUAD 中各种细胞类型的 UMAP 图; C, DSG2 在 LUAD 各种细胞类型上的分布情况; D, DSP 在 LUAD 各种细胞类型上的分布情况; E, PLOD2 在 LUAD 各种细胞类型上的分布情况; F, CAV1 在 LUAD 各种细胞类型上的分布情况; G, MYH11 在 LUAD 各种细胞类型上的分布情况。

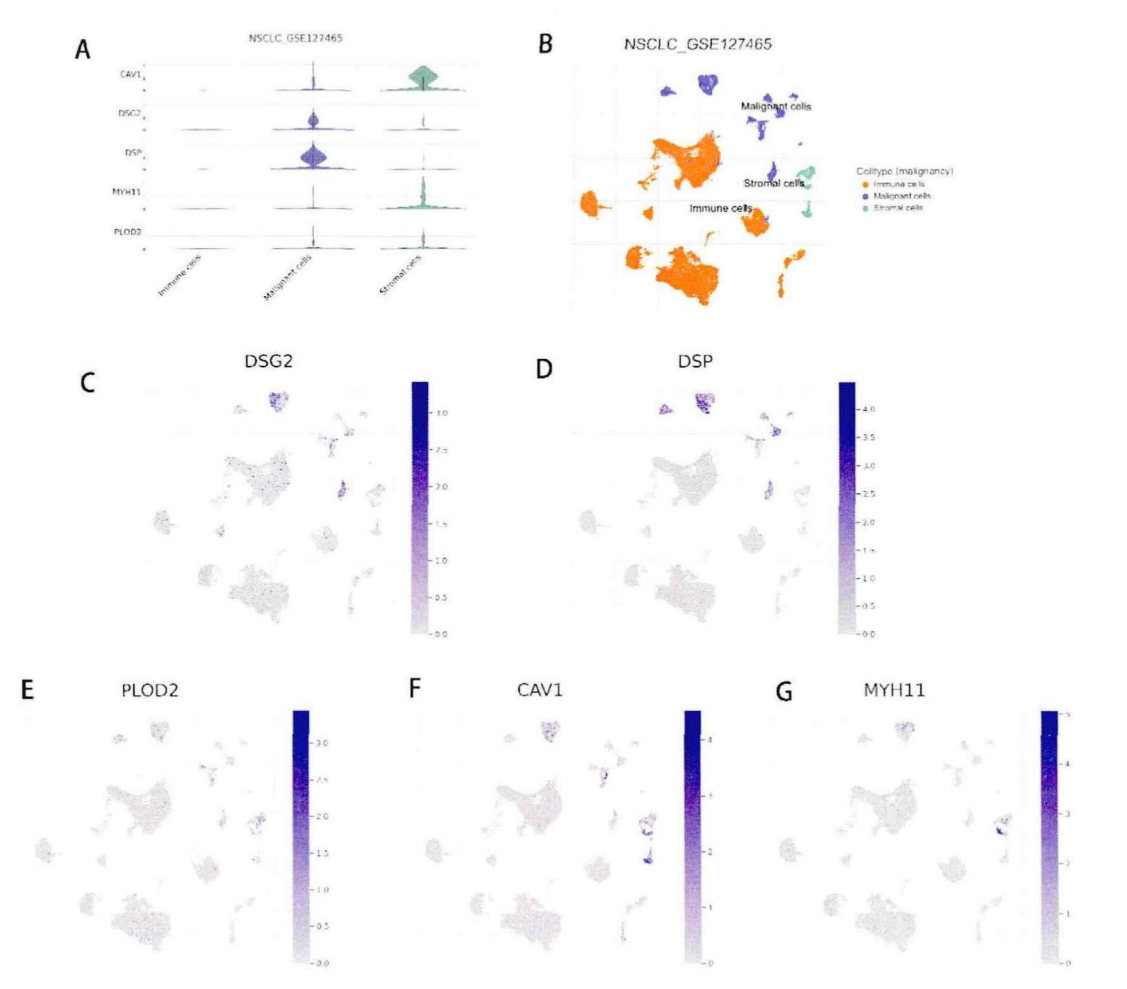


图 20. GSE127465 数据集中 MiAg DEGs 表达和分布的情况

A, MiAg DEGs 在 LUAD 三种主要类型细胞上的表达情况；B, LUAD 中各种细胞类型的 UMAP 图；C, DSG2 在 LUAD 各种细胞类型上的分布情况；D, DSP 在 LUAD 各种细胞类型上的分布情况；E, PLOD2 在 LUAD 各种细胞类型上的分布情况；F, CAV1 在 LUAD 各种细胞类型上的分布情况；G, MYH11 在 LUAD 各种细胞类型上的分布情况。

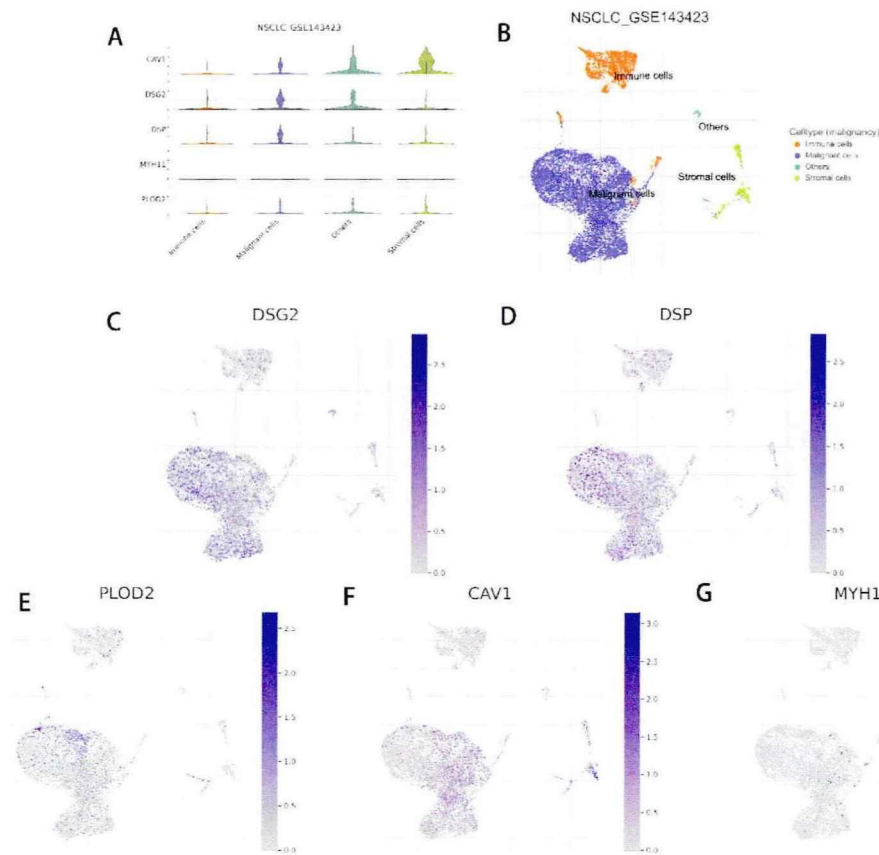


图 21. GSE143423 数据集中 MiAg DEGs 表达和分布的情况

A, MiAg DEGs 在 LUAD 四种主要类型细胞上的表达情况; B, LUAD 中各种细胞类型的 UMAP 图; C, DSG2 在 LUAD 各种细胞类型上的分布情况; D, DSP 在 LUAD 各种细胞类型上的分布情况; E, PLOD2 在 LUAD 各种细胞类型上的分布情况; F, CAV1 在 LUAD 各种细胞类型上的分布情况; G, MYH11 在 LUAD 各种细胞类型上的分布情况。

十五, MiAg DEGs 相关蛋白在 LUAD 中的定位表达情况及临床相关性

我们从 2018 年 1 月-12 月就诊于中国医学科学院肿瘤医院胸外科并接受手术治疗的 96 例 LUAD 患者的手术石蜡标本, 并制成石蜡切片。应用多重荧光免疫组织化学染色, 将荧光染料分别标记 PLOD2 (橙色)、DSG2 (绿色)、DSP (红色)、CAV1 (青色) 及 MYH11 (黄色), 应用 DAPI 标记细胞核 (图 22)。结果显示, 在 LUAD 细胞中存在 MiAg DEGs 相关蛋白的表达 (图 22A)。其中 DSG2 (图 22B) 和 DSP (图 22C) 主要在 LUAD 细胞中的定位和表达。PLOD2 (图 22C) 在

LUAD 细胞中少量定位和表达。而 CAV1 (图 22D) 和 MYH11 (图 22E) 在 LUAD 细胞中几乎不表达。

随后我们应用 Image-Pro Plus 6.0 软件对 96 例患者的多重荧光免疫组织化学染色图像进行分析, 同时结合患者的临床特征 (年龄、性别、肿瘤病理分期、肿瘤分化程度、生存状态), 进一步分析 MiAg DEGs 对 LUAD 患者的影响, 结果显示, 在年龄 < 60 岁和年龄 ≥ 60 岁 LUAD 患者中, 5 个 MiAg DEGs 在两组患者中均未显示统计学差异 (图 23A)。同样, 在不同的性别的 LUAD 患者中, 5 个 MiAg DEGs 在两组患者中也未显示统计学差异 (图 23B)。在 LUAD 不同的肿瘤分期中, DSG2 在 LUAD I 期和 III 期中表达具有显著差异 (图 24A); DSP 在 LUAD I 期中表达明显升高, 在 I 期和 II 期以及 I 期和 III 期中的表达具有统计学差异 (图 24B); PLOD2、CAV1 和 MYH11 在 LUAD 不同的肿瘤分期中的表达无明显统计学差异 (图 24C-E)。在肿瘤的不同分化水平中, DSG2 在 LUAD 低分化组中显著升高, 且在高分化和低分化组以及中分化和低分化组中的表达具有显著差异 (图 24F); DSP 在 LUAD 低分化组中表达降低, 在高分化组中表达升高, 且在高分化和低分化组中的表达具有统计学差异 (图 24G); PLOD2、CAV1 和 MYH11 在 LUAD 不同的分化分组中的表达无明显统计学差异 (图 24H-J)。在 LUAD 患者的不同生存状态中, DSG2 在死亡组中有较明显的升高, 而 DSP 在生存患者中表达显著升高 (图 24K; 图 24L); PLOD2、CAV1 和 MYH11 在 LUAD 患者的不同生存状态中的表达无明显统计学差异 (图 24M-O)。

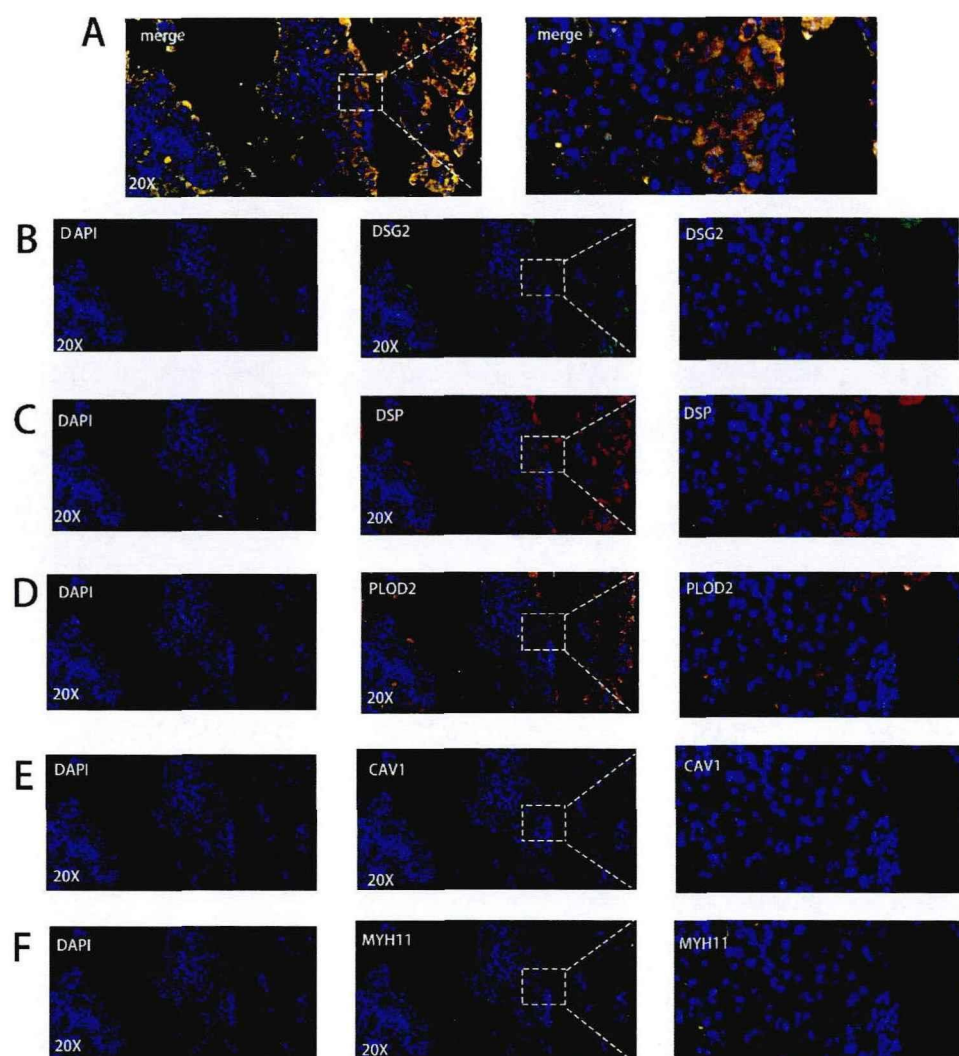


图 22. MiAg DEGs 相关蛋白在 LUAD 中的定位表达情况

A, merge 图像显示在 LUAD 细胞中存在 MiAg DEGs 相关蛋白的表达; B, 细胞核 DAPI 染色以及 DSG2 在 LUAD 细胞上的表达和定位; C, 细胞核 DAPI 染色以及 DSP 在 LUAD 细胞上的表达和定位; D, 细胞核 DAPI 染色以及 PLOD2 在 LUAD 细胞上的表达和定位; E, 细胞核 DAPI 染色以及 CAV1 在 LUAD 细胞上的表达和定位; F, 细胞核 DAPI 染色以及 MYH11 在 LUAD 细胞上的表达和定位。

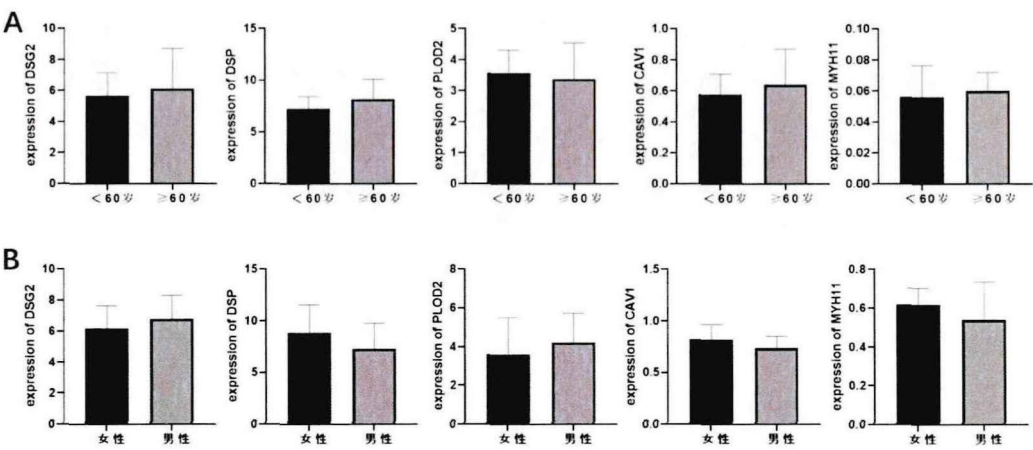


图 23. LUAD 中不同年龄和性别因素中 MiAg DEGs 的表达变化
A，年龄<60 岁和年龄≥60 岁 LUAD 患者中 MiAg DEGs 的表达差异。B，不同性别 LUAD 患者中 MiAg DEGs 的表达差异。

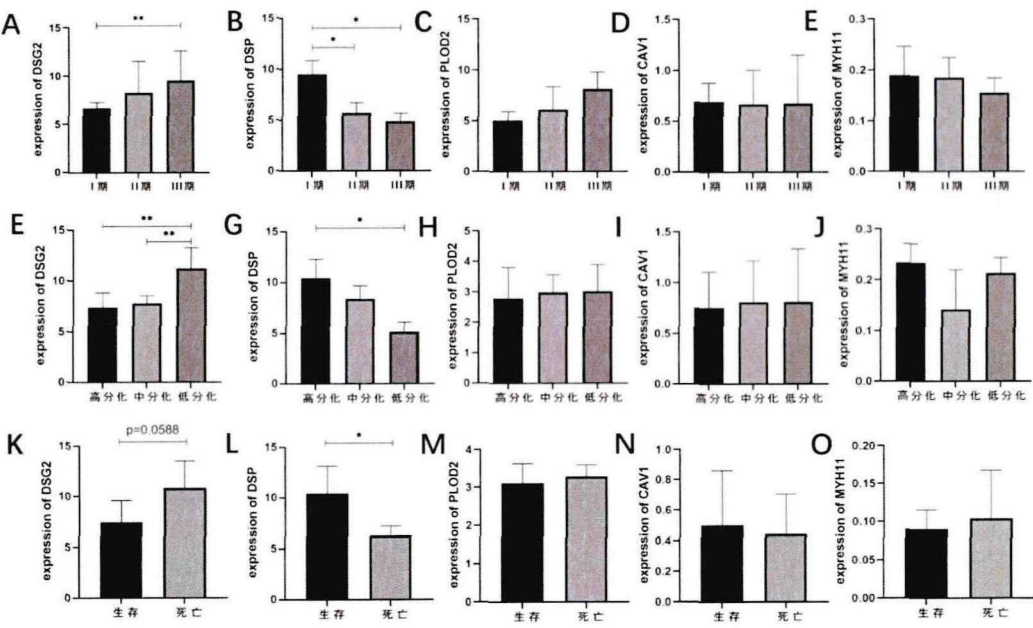


图 24. LUAD 中不同临床特征中 MiAg DEGs 表达变化
A-E，不同病理分期的 LUAD 患者中 MiAg DEGs 表达变化；F-J，不同肿瘤分化水平的 LUAD 患者中 MiAg DEGs 表达变化；K-O，LUAD 患者不同生存状态下 MiAg DEGs 表达变化。（*, $P<0.05$; **, $P<0.01$ ）

十六，DSP 在 LUAD 细胞中表达下调

我们利用 Real-Time qPCR 检测了 DSP 在正常人支气管上皮细胞系 BEAS-2B、人 LUAD 细胞系 (H1395、A549 及 H1975) 和人非小细胞肺癌细胞系 (H157) 中的表达情况, 结果显示, 在 LUAD 细胞系和非小细胞肺癌细胞系中, DSP 的表达水平明显低于正常人支气管上皮细胞系 BEAS-2B (图 25)。(*, $P<0.05$; ***, $P<0.001$;))

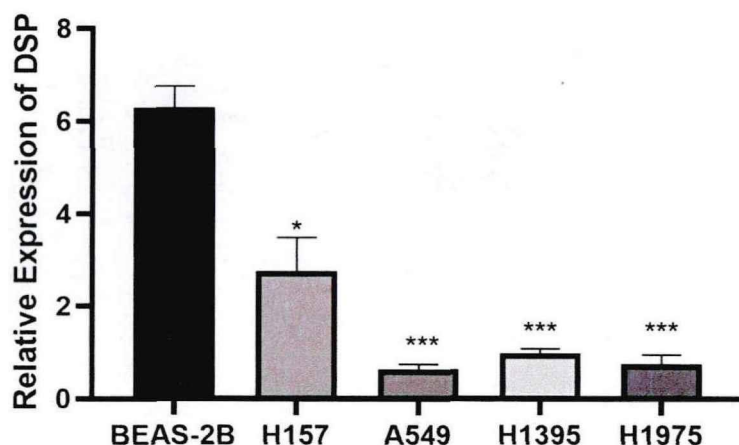


图 25. DSP 在 LUAD 细胞系中表达降低
(*, $P<0.05$; ***, $P<0.001$;))

十七, DSP 在 LUAD 细胞中过表达和敲减效果验证

为了进一步研究外源性 DSP 在 LUAD 中的作用, 我们选取了 DSP 表达较低三个 LUAD 细胞系 (H1975、A549 和 H1395), 并对其进行了 DSP 的稳定过表达。结果显示, 与对照组相比, 过表达的细胞系 H1975 (图 26A-C)、A549 (图 26D-F) 及 H1395 (图 26G-I) 中, DSP 的表达显著升高。接下来, 为了评价内源性 DSP 对 LUAD 的作用, 我们在 H1975、A549 及 H1395 这三种细胞系中利用 siRNA 对 DSP 进行敲减。验证结果显示, 与对照组相比 (NC-siRNA), 敲减组 (siRNA) DSP 的表达水平明显降低。

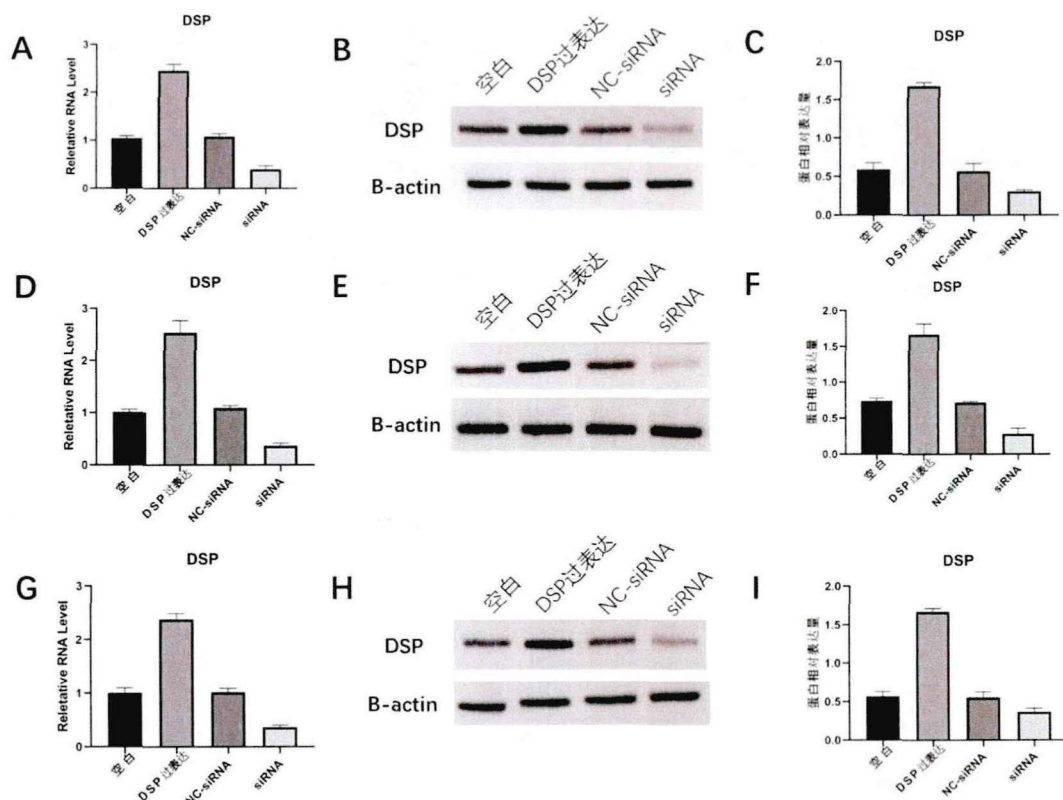


图 26. DSP 在 LUAD 细胞中过表达和敲减验证

A-C, H1975 细胞系中 DSP 转染细胞模型构建并在 RT-PCR (A) 和 western blot (B 和 C) 中验证。D-F, A549 细胞系中 DSP 转染细胞模型构建并在 RT-PCR (D) 和 western blot (E 和 F) 中验证。G-I, H1395 细胞系中 DSP 转染细胞模型构建并在 Real Time qPCR (G) 和 western blot (H 和 I) 中验证。

十八, DSP 抑制 LUAD 细胞的增殖

平板克隆实验结果显示, 在 H1975、A549 及 H1395 细胞系中, DSP-siRNA 组的克隆形成能力明显高于 NC-siRNA 组 (图 27A、27C 和 27E)。DSP 过表达组的细胞克隆形成能力明显低于对照组 (图 27A、27C 和 27E)。CCK-8 增殖实验结果显示, 48h 和 72h 时, 在 H1975、A549 及 H1395 细胞系中, DSP-siRNA 组的细胞活力明显高于 NC-siRNA 组。DSP 过表达组的细胞活力明显低于对照组 (图 27B、27D 和 27F)。

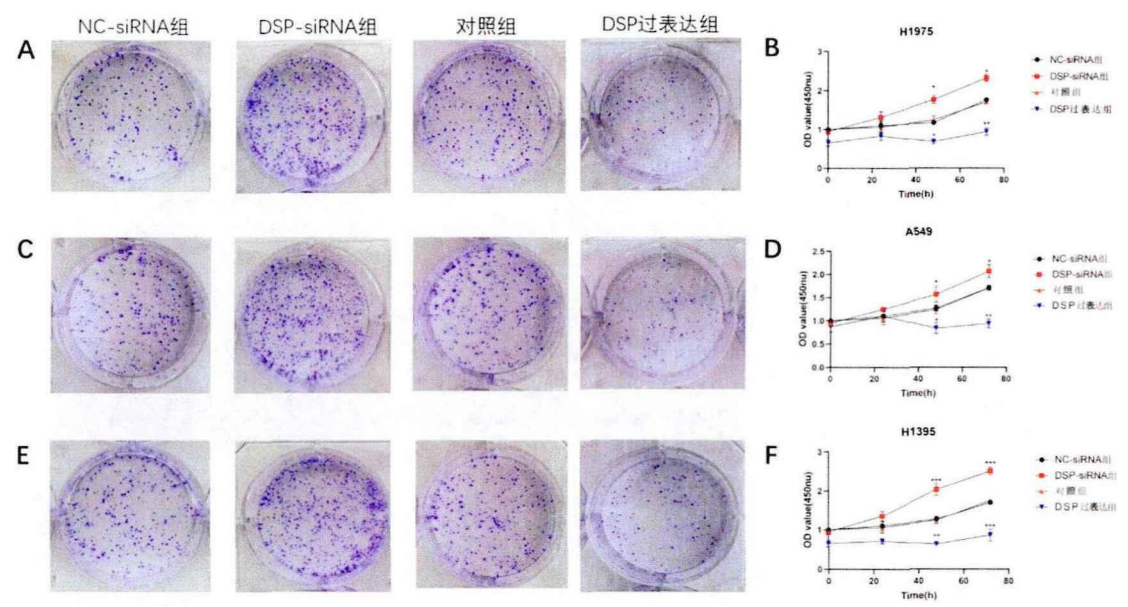


图 27. DSP 对 LUAD 细胞增殖能力的影响

A-B, 在 H1975 细胞中, 平板克隆实验 (A) 检测细胞克隆形成能力和 CCK-8 细胞增殖实验 (B) 检测不同时间点的细胞活力; C-D, 在 A549 细胞中, 平板克隆实验 (C) 检测细胞克隆形成能力和 CCK-8 细胞增殖实验 (D) 检测不同时间点的细胞活力; E-F, 在 H1395 细胞中, 平板克隆实验 (E) 检测细胞克隆形成能力和 CCK-8 细胞增殖实验 (F) 检测不同时间点的细胞活力。(*, $P<0.05$; **, $p<0.01$; ***, $P<0.001$)

十九, DSP 促进 LUAD 细胞的凋亡

流式细胞学检测结果显示, 在 H1975 细胞中, 过表达组的细胞凋亡率显著高于对照组; DSP-siRNA 组的细胞凋亡率明显低于 NC-siRNA 组 (图 28A 和图 28B)。在 A549 细胞中, 过表达组的细胞凋亡率显著高于对照组; DSP-siRNA 组的细胞凋亡率明显低于 NC-siRNA 组 (图 28C 和图 28D)。在 H1395 细胞中, 过表达组的细胞凋亡率显著高于对照组; DSP-siRNA 组的细胞凋亡率明显低于 NC-siRNA 组 (图 28E 和图 28F)。

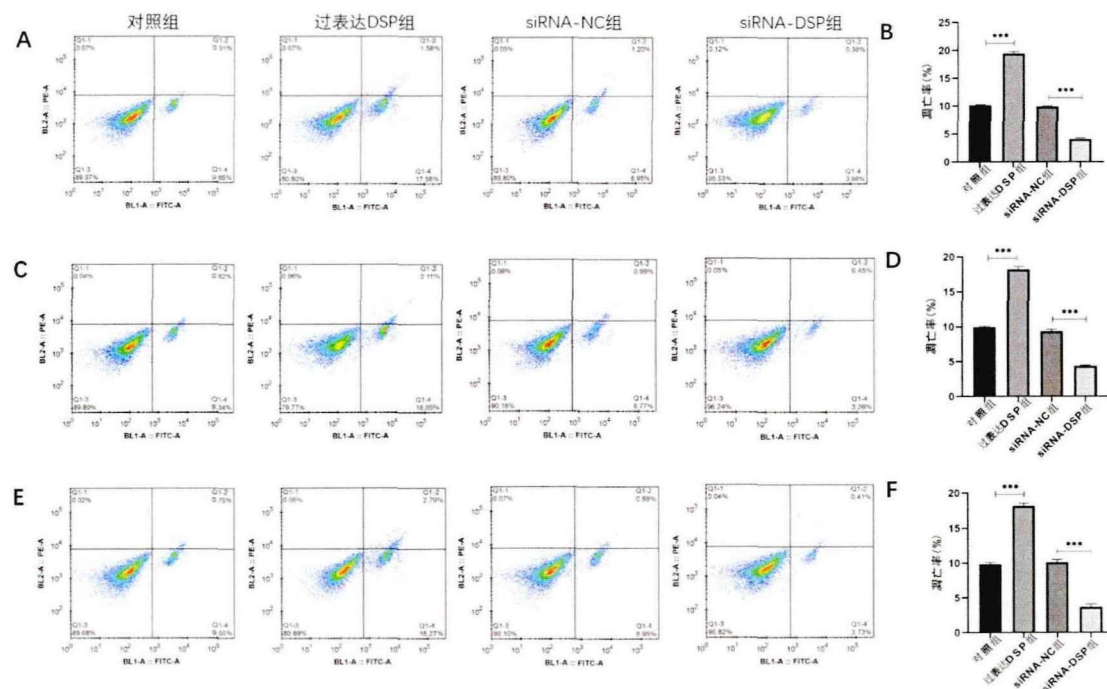


图 28. DSP 对 LUAD 细胞凋亡的影响

A-B, 在 H1975 细胞系中, DSP 对细胞凋亡的影响; C-D, 在 A549 细胞系中, DSP 对细胞凋亡的影响; E-F, 在 H1395 细胞系中, DSP 对细胞凋亡的影响。

二十, DSP 抑制 LUAD 细胞的迁移和侵袭

划痕实验结果显示, 在 H1975 细胞系中, DSP 过表达组的细胞迁移率明显低于对照组; DSP-siRNA 组的细胞凋亡率明显高于 NC-siRNA 组 (图 29); 在 A549 细胞系中, DSP 过表达组的细胞迁移率明显低于对照组; DSP-siRNA 组的细胞凋亡率明显高于 NC-siRNA 组 (图 30); 在 H1395 细胞系中, DSP 过表达组的细胞迁移率明显低于对照组; DSP-siRNA 组的细胞凋亡率明显高于 NC-siRNA 组 (图 31)。

Transwell 实验结果显示, 在 H1975 细胞系中, DSP 过表达组细胞迁移能力显著下降; DSP-siRNA 组的细胞迁移能力显著高于 NC-siRNA 组; 在 A549 细胞系中, DSP 过表达组的细胞迁移能力明显低于对照组; DSP-siRNA 组的细胞迁移能力明显高于 NC-siRNA 组; 在 H1395 细胞系中, DSP 过表达组的细胞迁移能力明显低于对照组; DSP-siRNA 组的细胞迁移能力明显高于 NC-siRNA 组。

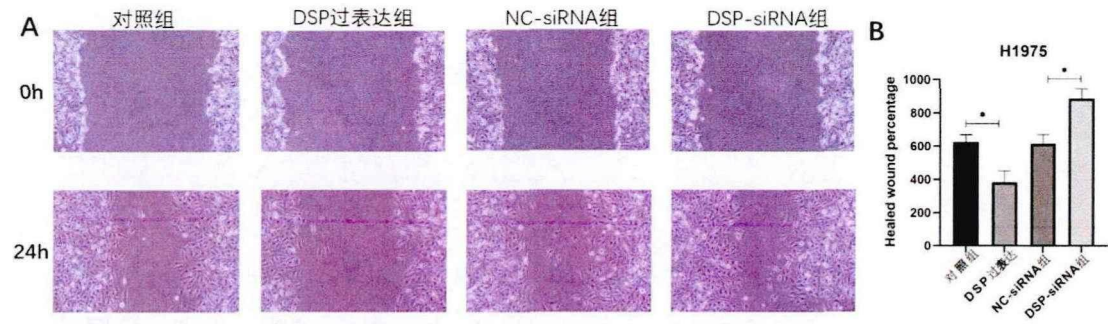


图 29. 划痕实验检测 DSP 对 H1975 细胞迁移的影响

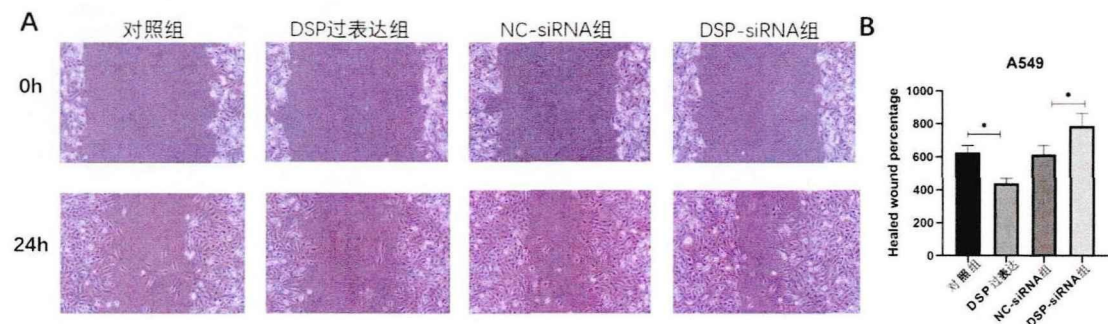


图 30. 划痕实验检测 DSP 对 A549 细胞迁移的影响

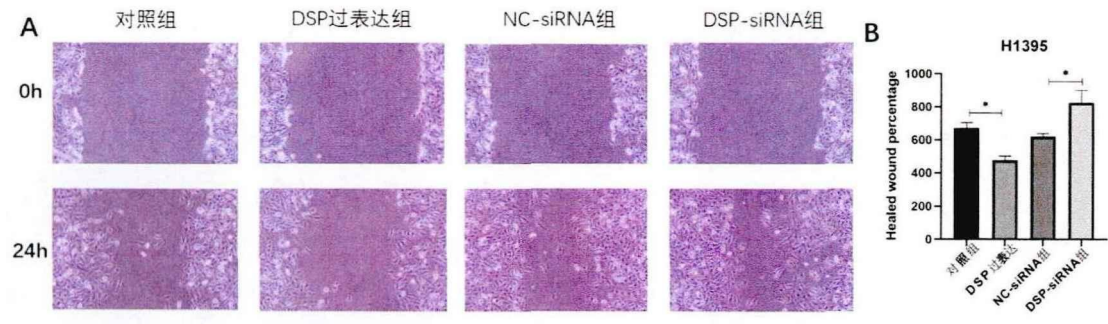


图 31. 划痕实验检测 DSP 对 H1395 细胞迁移的影响

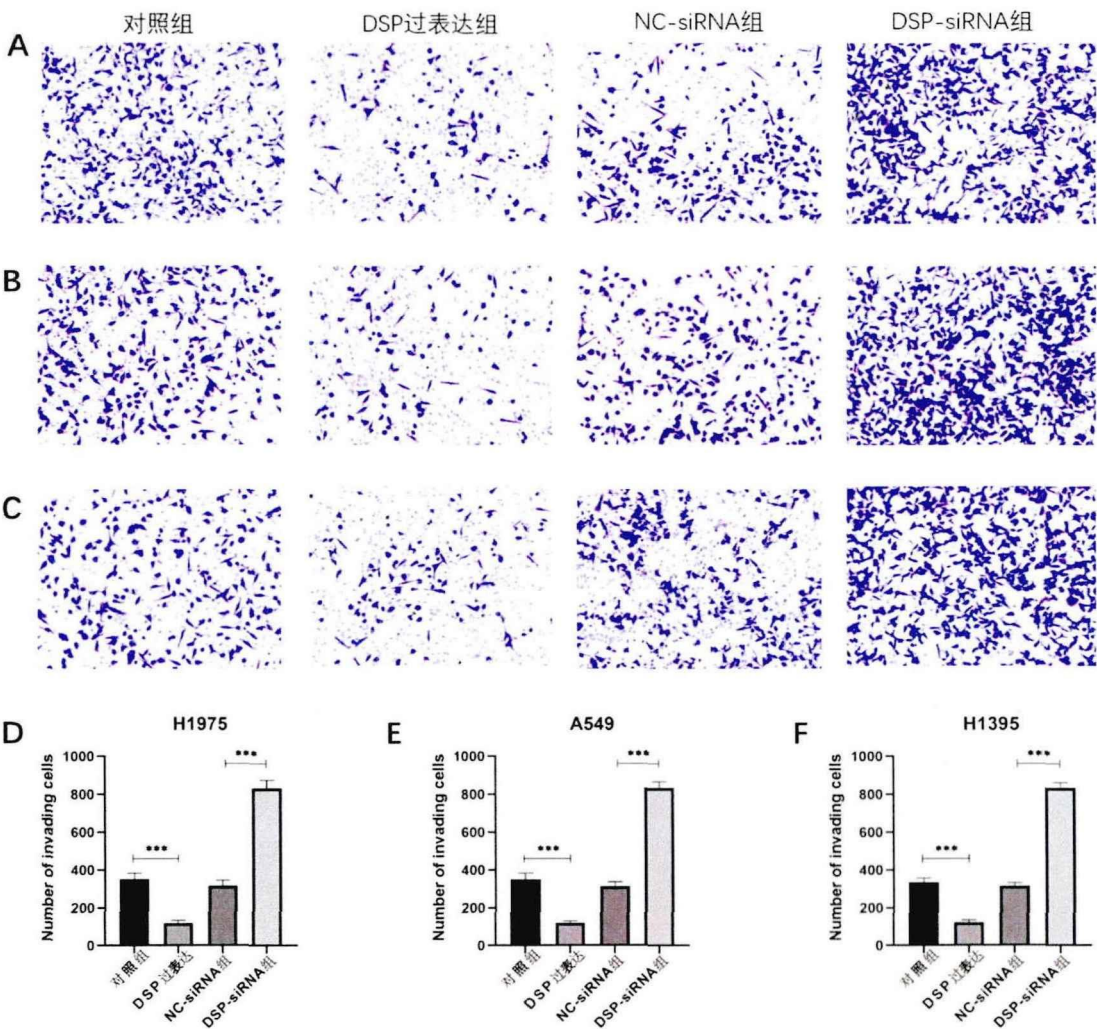


图 32. DSP 对 LUAD 细胞侵袭能力的影响

Transwell 实验检测 DSP 对 H1975 细胞（A 和 D）、A549 细胞（B 和 E）和 H1395 细胞（C 和 F）侵袭能力的影响。

讨论

作为非小细胞肺癌最常见的类型之一，LUAD 对患者生命构成了巨大威胁。研究显示，晚期 LUAD 患者的 5 年生存率不超过 20%[Siegel et al, 2012]。尽管放疗、化疗、靶向治疗等医疗技术不断进步，但大多数患者的生存率改善有限[Zhou et al, 2018]。因此，如果能够构建出预测 LUAD 患者的预后模型并找到影响 LUAD 预后的重要分子，将有助于 LUAD 患者治疗决策的选择。

研究表明，线粒体自噬不仅参与维持细胞的稳态和存活，还与包括癌症在内的多种疾病的发生有关。衰老是一个不可避免的生物学过程，以往的研究表明，伴随着细胞功能的逐渐衰退，衰老与癌症发病率增加之间存在密切联系。然而，迄今为止，线粒体自噬和衰老相关的共表达基因在 LUAD 的发生发展、治疗和预后中的作用尚未见报道。本研究发现，线粒体自噬和衰老相关的 5 个基因（CAV1、DSG2、DSP、MYH11、PLOD2）在 LUAD 中发挥着重要作用。这一发现拓展了我们对线粒体自噬、衰老过程和 LUAD 之间错综复杂的相关性的理解，从而为即将进行治疗和干预提供了一个潜在的方向。

通过对 TCGA-LUAD、GSE10072 和 GSE30219 数据集的分析表明，5 个 MiAg DEGs 的表达水平在 LUAD 和正常组织之间存在显著差异。结果显示，与正常肺组织相比，CAV1 和 MYH11 在 LUAD 中的表达显著降低，而其他三个基因（DSG2、DSP、PLOD2）在 LUAD 中的表达升高。

在 5 个 MiAg DEGs 中，CAV1 是许多信号分子的负调控因子，它的失调、缺失或突变与包括癌症在内的多种疾病的病理特征相关。研究发现，将 PAM212 细胞中的 CAV1 敲除会促进肿瘤的生长，并增强肿瘤在体外和体内的侵袭能力。具体机制可能与 ERK1/2 和 MAPK 通路过度激活有关[Trimmer et al, 2013]。既往研究发现，MYH11 基因编码的肌球蛋白重链是一种属于肌球蛋白重链家族的平滑肌肌球蛋白[Matsuoka et al, 1993]。MYH11 基因突变还与多种肿瘤相关，包括结直肠癌[Wang et al, 2014]、膀胱癌[Hu et al, 2019]、喉癌[Su et al, 2017]以及头颈部癌症[Islam et al, 2018]。这可能是因为 MYH11 在参与细胞迁移、与细胞粘附蛋白相互作用以及在抑制肿瘤生长方面的潜在作用[Pollard et al, 1974]。此外，研究还证实了 MYH11 在非小细胞肺癌中的关键作用[Ma et al, 2019]。

Desmoglein-2 (DSG2) 是钙粘蛋白家族中一种重要的跨膜糖蛋白，通过在细胞之间形成桥粒连接的复合物来介导细胞间粘附和信号传导[Johnson et al, 2014]。DSG2 在非小细胞肺癌中表达较高，敲除 DSG2 可抑制非小细胞肺癌的进展[Cai et al, 2017]。Katharina 等人[Hutz et al, 2017]发现了一种新的胰腺癌细胞迁移途径，其中 DSG2 的缺失会通过失调的 MAPK 信号转导发挥作用。既往研究表明，DSG2 通

过 EGFR 信号通路在 LUAD 的发展和进展中具有独特作用,结果显示,DSG2 在细胞膜上直接与 EGFR 相互作用,并通过 EGFR-Src-Rac1-PAK1 信号通路调节肿瘤进展[Jin et al, 2020]。这个研究结果也与我们研究所得到的结果相吻合。

Desmoplakin (DSP) 是桥粒的关键成分,桥粒对细胞-细胞粘附很重要,但该蛋白也被证明会影响细胞增殖、分化、迁移和凋亡[Serrano-Gomez et al, 2016]。还有研究表明,支气管细胞中 DSP 的缺失导致上皮-间充质转化和细胞外基质基因的表达增加,从而使细胞迁移增加[Pascual et al, 2017]。除了维持桥粒的完整性外, DSP 还通过调节癌症细胞中的各种信号通路,作为肿瘤抑制因子发挥重要作用[Chun et al, 2010; Yang et al, 2012]。然而到目前为止, DSP 在 LUAD 中的作用以及对 LUAD 细胞的影响目前尚不清楚。我们的研究结果显示, DSP 可以抑制 LUAD 细胞的侵袭和迁移,并对细胞增殖有显著的抑制作用,同时促进细胞的凋亡,从而证明 DSP 在 LUAD 中发挥负向调控作用,对 LUAD 细胞的生长和转移具有抑制作用。但在对 LUAD 患者的分析中我们发现,在 LUAD 早期以及高分化的 LUAD 细胞中, DSP 的表达明显高于晚期以及低分化的 LUAD,这也印证了细胞水平验证上得到的证据。

Procollagen-lysine, 2-oxoglutarate 5-dioxygenase 2 (PLOD2) 是 PLOD 家族的成员,是介导稳定的胶原交联形成的关键基因。PLOD2 可能在癌症转移中发挥重要作用。先前的研究表明, PLOD2 在各种癌症类型中高表达,包括肝细胞癌、口腔鳞状细胞癌等[Ueki et al, 2020]。此外, PLOD2 的表达与癌症患者的预后有关[Shao et al, 2022]。在我们的研究中, PLOD2 也表达在 LUAD 细胞中,但在临床相关因素的分析中并未发现其显著的统计学改变。因此, PLOD2 在 LUAD 中的作用还有待于进一步研究。

此外,我们还对 MiAg DEGs 进行了 GO 和 GSEA 分析,结果显示,在 TCGA-LUAD 数据集中, MiAg scores 的高和低的在激活的 MAPK 通路中显著富集,而且还与 DNA 损伤检查点、NOTCH 通路、NF-KB 等通路有关。研究发现, MAPK 通路在多种细胞功能,尤其是细胞增殖、分化和凋亡中具有调节作用[Choi et al, 2003],而 MAPK 信号通路也是导致 LUAD 和其他肿瘤发生的原因之一[Stutvoet et al, 2019]。在 LUAD 中, MAPK 通路的活性参与了 EGF 和 IFN- γ 诱导的 PD-L1 的表达,而没有靶向基因的改变,这可能是提高免疫疗法疗效的一个靶点[Stutvoet et al, 2019]。

此外,我们的研究还发现了 MiAg-RGs 在 LUAD 免疫治疗中的作用。近年来,免疫疗法的最新进展彻底改变了癌症治疗的格局,这也凸显了肿瘤免疫微环境的重要性。我们的研究结果表明, TCGA-LUAD 数据集中的高低 MiAg scores 组在 IPS 上表现出显著差异 ($P < 0.001$),另一方面,在 TCGA-LUAD 数据集中,两组 MiAg scores 之间的 IPS-CTLA4 差异显著 ($P < 0.001$)。既往研究发现, Cav-1 (-/-) 供体 T 细胞会产生更多的 Tregs[Schonle et al, 2016]。在我们的研究中也观察到了类似的

相关性。此外,我们的研究结果表明,在 MiAg scores 低分组中,CAV-1 和 Tregs 之间呈负相关。我们的研究揭示了一项新发现,即在 MiAg scores 高分组中,Tregs 与 MYH11、CYCS 之间存在负相关。CD4 记忆激活的 T 细胞与 NME1/6/7 之间存在明显的正相关。综上,我们的研究表明,MiAg DEGs 可能有助于 LUAD 的免疫治疗,并可能为 LUAD 的免疫治疗提供新的线索。

在之前的研究中,研究人员检测了 27 个线粒体自噬相关基因[Dai et al, 2022]和 21 个衰老相关基因[Zhang et al, 2022]对 LUAD 的预后和免疫浸润影响。然而,之前的研究并未分析与这两种重要表型相关的基因。在我们的研究中,我们在 TCGA-LUAD 数据集中探讨了 MiAg DEGs 的预后价值。通过 Lasso 回归,我们最终分别用 CAV1、DSG2、DSP、MYH11、PLOD2 这 5 个基因构建了 MiAg DEGs 的预后模型,结果显示,在 TCGA-LUAD 数据集中,MiAg DEG 预后模型的 RiskScore 对 LUAD 患者的 OS 具有很高的预测准确性($P<0.001$)。同时,我们还分析了 MiAg DEGs 预后模型在 GSE3141 数据集中的预测准确性,结果显示,MiAg DEGs 预后模型的 RiskScore 对 GSE3141 数据集中 LUAD 患者 OS 的预测具有一定的准确性($AUC1=0.706$, $AUC3=0.752$, $AUC5=0.692$)。上述结果表明,我们基于 MiAg DEGs 的预后模型可以预测 LUAD 患者的预后。与之前的研究相比,我们的研究同时考虑了线粒体自噬和衰老两个表型,涵盖的基因更少,更有利于后续的推广和应用验证。为了进一步验证我们的 MiAg scores,我们采用单变量和多变量 Cox 回归分析了 TCGA-LUAD 数据集中 MiAg scores 风险分数和五个临床变量(T 分期、N 分期、M 分期、年龄和性别)与 LUAD 患者临床结局的关联,结果显示 T 分期和 MiAg scores 在多变量 Cox 回归模型中的效用明显高于其他变量。同时,我们还评估了多变量 Cox 回归模型的诺模图,发现我们构建的多变量回归模型的临床预测效果为 5 年>3 年>1 年。在我们的研究中,我们证实了这 5 个 MiAg DEGs 与 LUAD 患者预后的相关性。

但该研究的结果还需进一步的研究验证,未来应在更大规模的独立队列中验证这一模型,并探索这些基因在 LUAD 中发挥作用背后的详细分子机制。了解这些机制将为开发有效治疗策略开辟新的途径。因此我们对后续的机制研究进行了初步的探索。首先我们在本中心的 LUAD 组织中,利用 Real Time qPCR 成功测定了 5 个目标基因在 LUAD 组织及癌旁组织中的相对表达量。结果显示,相较于癌旁组织,LUAD 组织中 DSG2 的表达水平显著上调($P<0.05$),而 CAV1、MYH11、PLOD2 以及 DSP 在 LUAD 组织和癌旁组织中的表达未见明显的统计学差异。接着我们利用公共数据库进行了单细胞测序分析,三个数据集的分析结果显示,在 LUAD 中,DSG2 和 DSP 主要分布和表达于恶性肿瘤细胞中。PLOD2 和 CAV1 虽然少量表达于恶性肿瘤细胞中,但主要分布于基质细胞和其他类型的细胞中。而 MYH11 则在

LUAD 主要的细胞类型中未见规律的表达和分布。然后我们也通过本中心 LUAD 患者的石蜡切片进行多重荧光免疫组织化学染色，结果提示 DSG2 和 DSP 主要分布和表达在 LUAD 细胞中，而且 DSG2 在 5 年后死亡、分期晚以及分化差的 LUAD 患者中表达升高，这与之前的研究结果相似[Jin et al, 2020]。由于 DSG2 基因对 LUAD 发生发展的影响已经得到了较为全面的验证，因此在我们的研究中，我们主要关注 DSP 基因对 LUAD 发生发展的影响。在我们的研究中，DSP 的分析结果显示，其在癌症分期早以及分化好的 LUAD 患者中表达升高，在癌症分期晚及分化差的 LUAD 患者中表达降低，这些结果提示 DSP 对 LUAD 细胞的发生发展可能产生负向的调控作用。初步的细胞功能实验结果显示，DSP 可以抑制 LUAD 细胞的侵袭和迁移，并对细胞增殖有显著的抑制作用，同时促进细胞的凋亡，这些结果与之前的推测相一致。

本研究虽然取得了一些成果，但也具有一定的局限性。首先，我们的研究是回顾性研究，仅依赖于公开的数据库和单中心的数据。此外，我们分析中使用的数据集可能存在批间差异。

小结

1, 本研究筛选了 LUAD 中与线粒体自噬和衰老相关的差异表达基因, 经过筛选, 5 个 MiAg DEGs (CAV1、DSG2、DSP、MYH11、PLOD2) 与 LUAD 患者的预后具有显著相关性。

2, 构建的 MiAg DEGs 预后模型在 LUAD 患者的预后中预测能力良好。

3, 5 个 MiAg DEGs 中 DSG2 和 DSP 主要分布和表达在 LUAD 细胞上, 既往研究已证实 DSG2 对 LUAD 细胞的增殖和转移具有促进作用。

4, DSP 对 LUAD 细胞的增殖、侵袭和迁移具有显著的抑制作用, 这提示 DSG2 和 DSP 可能在 LUAD 的发生发展中发挥作用并可能成为潜在的治疗靶点。

参考文献

- Athar A et al. (2019). ArrayExpress update - from bulk to single-cell expression data. *Nucleic Acids Res.* 47(D1):D711-D715.
- Azizi E et al. (2018). Single-Cell Map of Diverse Immune Phenotypes in the Breast Tumor Microenvironment. *Cell.* 174(5):1293-1308 e36.
- Barrett T et al. (2007). NCBI GEO: mining tens of millions of expression profiles--database and tools update. *Nucleic Acids Res.* 35(Database issue):D760-5.
- Barrett T et al. (2013). NCBI GEO: archive for functional genomics data sets--update. *Nucleic Acids Res.* 41(Database issue):D991-5.
- Becht E et al. (2018). Dimensionality reduction for visualizing single-cell data using UMAP. *Nat Biotechnol.*
- Bild A H et al. (2006). Oncogenic pathway signatures in human cancers as a guide to targeted therapies. *Nature.* 439(7074):353-7.
- Bray F et al. (2018). Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin.* 68(6):394-424.
- Cai F et al. (2017). Desmoglein-2 is overexpressed in non-small cell lung cancer tissues and its knockdown suppresses NSCLC growth by regulation of p27 and CDK2. *J Cancer Res Clin Oncol.* 143(1):59-69.
- Chen B et al. (2018). Profiling Tumor Infiltrating Immune Cells with CIBERSORT. *Methods Mol Biol.* 1711:243-259.
- Chen H et al. (2020). Age-related mutational signature negatively associated with immune activity and survival outcome in triple-negative breast cancer. *Oncoimmunology.* 9(1):1788252.
- Chen H et al. (2021). Development and Validation of a Novel Mitophagy-Related Gene Prognostic Signature for Hepatocellular Carcinoma Based on Immunoscore Classification of Tumor. *J Oncol.* 2021:5070099.
- Choi K C et al. (2003). Mitogen-activated protein kinases in normal and (pre)neoplastic ovarian surface epithelium. *Reprod Biol Endocrinol.* 1:71.
- Chong W et al. (2021). Association of clock-like mutational signature with immune checkpoint inhibitor outcome in patients with melanoma and NSCLC. *Mol Ther Nucleic Acids.* 23:89-100.
- Chun M G et al. (2010). Genetic deletion of the desmosomal component desmoplakin promotes tumor microinvasion in a mouse model of pancreatic neuroendocrine carcinogenesis. *PLoS Genet.* 6(9):e1001120.
- Colaprico A et al. (2016). TCGAAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data. *Nucleic Acids Res.* 44(8):e71.
- Dai D et al. (2022). A Comprehensive Analysis of the Effects of Key Mitophagy Genes on the Progression and Prognosis of Lung Adenocarcinoma. *Cancers (Basel).* 15(1).
- Dai K et al. (2019). PINK1 depletion sensitizes non-small cell lung cancer to glycolytic inhibitor 3-bromopyruvate: Involvement of ROS and mitophagy. *Pharmacol Rep.* 71(6):1184-1189.
- Davis S et al. (2007). GEOquery: a bridge between the Gene Expression Omnibus (GEO) and BioConductor. *Bioinformatics.* 23(14):1846-7.
- Denisenko T V et al. (2021). Mitophagy in carcinogenesis and cancer treatment. *Discov Oncol.* 12(1):58.

- Ettinger D S et al. (2015). Non-Small Cell Lung Cancer, Version 6.2015. *J Natl Compr Canc Netw*. 13(5):515-24.
- Goldman M J et al. (2020). Visualizing and interpreting cancer genomics data via the Xena platform. *Nat Biotechnol*. 38(6):675-678.
- Guo J Y et al. (2011). Activated Ras requires autophagy to maintain oxidative metabolism and tumorigenesis. *Genes Dev*. 25(5):460-70.
- Hanzelmann S et al. (2013). GSEA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinformatics*. 14:7.
- He J et al. (2022). Identification and Validation of Aging-Related Genes in Idiopathic Pulmonary Fibrosis. *Front Genet*. 13:780010.
- Hu J et al. (2019). The identification of new biomarkers for bladder cancer: A study based on TCGA and GEO datasets. *J Cell Physiol*. 234(9):15607-15618.
- Hutz K et al. (2017). Loss of desmoglein 2 promotes tumorigenic behavior in pancreatic cancer cells. *Mol Carcinog*. 56(8):1884-1895.
- Islam T et al. (2018). Drug Targeting and Biomarkers in Head and Neck Cancers: Insights from Systems Biology Analyses. *OMICS*. 22(6):422-436.
- Jin R et al. (2020). Desmoglein-2 modulates tumor progression and osimertinib drug resistance through the EGFR/Src/PAK1 pathway in lung adenocarcinoma. *Cancer Lett*. 483:46-58.
- Johnson J L et al. (2014). Desmosomes: regulators of cellular signaling and adhesion in epidermal health and disease. *Cold Spring Harb Perspect Med*. 4(11):a015297.
- Lambrechts D et al. (2018). Phenotype molding of stromal cells in the lung tumor microenvironment. *Nat Med*. 24(8):1277-1289.
- Landi M T et al. (2008). Gene expression signature of cigarette smoking and its role in lung adenocarcinoma development and survival. *PLoS One*. 3(2):e1651.
- Liberzon A et al. (2015). The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst*. 1(6):417-425.
- Ma Q et al. (2019). Identification and validation of key genes associated with non-small-cell lung cancer. *J Cell Physiol*. 234(12):22742-22752.
- Matsuoka R et al. (1993). Human smooth muscle myosin heavy chain gene mapped to chromosomal region 16q12. *Am J Med Genet*. 46(1):61-7.
- Newman A M et al. (2015). Robust enumeration of cell subsets from tissue expression profiles. *Nat Methods*. 12(5):453-7.
- Park S Y (2018). Nomogram: An analogue tool to deliver digital knowledge. *J Thorac Cardiovasc Surg*. 155(4):1793.
- Pascual G et al. (2017). Targeting metastasis-initiating cells through the fatty acid receptor CD36. *Nature*. 541(7635):41-45.
- Patel A P et al. (2014). Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. *Science*. 344(6190):1396-401.
- Perkins N J et al. (2019). Combining Biomarker Calibration Data to Reduce Measurement Error. *Epidemiology*. 30 Suppl 2(Suppl 2):S3-S9.
- Peters M J et al. (2015). The transcriptional landscape of age in human peripheral blood. *Nat Commun*. 6:8570.
- Pollard T D et al. (1974). Actin and myosin and cell movement. *CRC Crit Rev Biochem*. 2(1):1-65.
- Puram S V et al. (2017). Single-Cell Transcriptomic Analysis of Primary and Metastatic Tumor

- Ecosystems in Head and Neck Cancer. *Cell*. 171(7):1611-1624 e24.
- Ritchie M E et al. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res*. 43(7):e47.
- Schonle A et al. (2016). Caveolin-1 regulates TCR signal strength and regulatory T-cell differentiation into alloreactive T cells. *Blood*. 127(15):1930-9.
- Serrano-Gomez S J et al. (2016). Regulation of epithelial-mesenchymal transition through epigenetic and post-translational modifications. *Mol Cancer*. 15:18.
- Shao Y et al. (2022). Proteomics profiling of colorectal cancer progression identifies PLOD2 as a potential therapeutic target. *Cancer Commun (Lond)*. 42(2):164-169.
- Sharma A et al. (2021). Mitochondrial dynamics and mitophagy in lung disorders. *Life Sci*. 284:119876.
- Siegel R et al. (2012). Cancer treatment and survivorship statistics, 2012. *CA Cancer J Clin*. 62(4):220-41.
- Stelzer G et al. (2016). The GeneCards Suite: From Gene Data Mining to Disease Genome Sequence Analyses. *Curr Protoc Bioinformatics*. 54:1 30 1-1 30 33.
- Stuart T et al. (2019). Comprehensive Integration of Single-Cell Data. *Cell*. 177(7):1888-1902 e21.
- Stutvoet T S et al. (2019). MAPK pathway activity plays a key role in PD-L1 expression of lung adenocarcinoma cells. *J Pathol*. 249(1):52-64.
- Su J et al. (2017). A recurrence model for laryngeal cancer based on SVM and gene function clustering. *Acta Otolaryngol*. 137(5):557-562.
- Subramanian A et al. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A*. 102(43):15545-50.
- Tataranni T et al. (2019). Dichloroacetate (DCA) and Cancer: An Overview towards Clinical Applications. *Oxid Med Cell Longev*. 2019:8201079.
- Trimmer C et al. (2013). Cav1 suppresses tumor growth and metastasis in a murine model of cutaneous SCC through modulation of MAPK/AP-1 activation. *Am J Pathol*. 182(3):992-1004.
- Ueki Y et al. (2020). PLOD2 Is Essential to Functional Activation of Integrin beta1 for Invasion/Metastasis in Head and Neck Squamous Cell Carcinomas. *iScience*. 23(2):100850.
- Van Calster B et al. (2018). Reporting and Interpreting Decision Curve Analysis: A Guide for Investigators. *Eur Urol*. 74(6):796-804.
- Wang C et al. (2020). Integrative analyses of single-cell transcriptome and regulome using MAESTRO. *Genome Biol*. 21(1):198.
- Wang H Q et al. (2011). Deregulated miR-155 promotes Fas-mediated apoptosis in human intervertebral disc degeneration by targeting FADD and caspase-3. *J Pathol*. 225(2):232-42.
- Wang R J et al. (2014). Down-regulated MYH11 expression correlates with poor prognosis in stage II and III colorectal cancer. *Asian Pac J Cancer Prev*. 15(17):7223-8.
- Wang Y et al. (2020). The Role of Mitochondrial Dynamics and Mitophagy in Carcinogenesis, Metastasis and Therapy. *Front Cell Dev Biol*. 8:413.
- Wang Y et al. (2021). Identification of HCC Subtypes With Different Prognosis and Metabolic Patterns Based on Mitophagy. *Front Cell Dev Biol*. 9:799507.
- Xu C et al. (2015). Identification of cell types from single-cell transcriptomes using a novel clustering method. *Bioinformatics*. 31(12):1974-80.
- Yang L et al. (2012). Desmoplakin acts as a tumor suppressor by inhibition of the Wnt/beta-catenin

- signaling pathway in human lung cancer. *Carcinogenesis*. 33(10):1863-70.
- Yost K E et al. (2019). Clonal replacement of tumor-specific T cells following PD-1 blockade. *Nat Med*. 25(8):1251-1259.
- Yu G et al. (2012). clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS*. 16(5):284-7.
- Yu G (2020). Gene Ontology Semantic Similarity Analysis Using GOSemSim. *Methods Mol Biol*. 2117:207-215.
- Zeng X et al. (2020). Poor glycemic control might compromise the efficacy of chemotherapy in non-small cell lung cancer patients with diabetes mellitus. *Cancer Med*. 9(3):902-911.
- Zhang L et al. (2018). Lineage tracking reveals dynamic relationships of T cells in colorectal cancer. *Nature*. 564(7735):268-272.
- Zhang L et al. (2020). Single-Cell Analyses Inform Mechanisms of Myeloid-Targeted Therapies in Colon Cancer. *Cell*. 181(2):442-459 e29.
- Zhang W et al. (2022). An aging-related signature predicts favorable outcome and immunogenicity in lung adenocarcinoma. *Cancer Sci*. 113(3):891-903.
- Zhou X et al. (2018). Analyzing EGFR mutations and their association with clinicopathological characteristics and prognosis of patients with lung adenocarcinoma. *Oncol Lett*. 16(1):362-370.
- Zhuo Z et al. (2021). Mitophagy-Related Gene Signature for Prediction Prognosis, Immune Scenery, Mutation, and Chemotherapy Response in Pancreatic Cancer. *Front Cell Dev Biol*. 9:802528.

肺腺癌预后模型的现状和研究进展 (文献综述)

[摘要] 肺腺癌作为一种常见的恶性肿瘤，其预后模型的研究对于指导临床治疗具有重要意义。近年来，随着生物信息学、高通量测序技术的发展以及人工智能算法的应用，肺腺癌预后模型的构建越来越注重多维度数据的整合，包括基因表达、蛋白质组学、临床特征及影像学资料。这些研究展示了通过深入分析肿瘤的分子特征和微环境因素，可以更准确地预测疾病的进展和患者的生存率。此外，个性化医疗的发展促进了预后模型向更精准、更具预测价值的方向进化。未来的研究需要在更大的患者群体中验证这些模型的临床适用性，并进一步探索模型背后肿瘤生物学的复杂性，以实现肺腺癌患者的精准治疗和管理。

[关键词] 肺腺癌；预后模型；基因；人工智能

肺癌，作为全球范围内最常见的恶性肿瘤之一，其发病率和死亡率持续居高不下[Skoulidis et al, 2019; Thai et al, 2021]。肺癌中约 80%-85%是非小细胞肺癌，而在非小细胞肺癌中，肺腺癌占了很大的比重[Siegel et al, 2021; Siegel et al, 2020]。根据世界卫生组织的统计，肺腺癌在全球肺癌中的比例逐年上升，已成为一个严重的公共卫生问题[Wadowska et al, 2020]。尽管在过去的几十年里，随着医疗技术的进步，包括手术[Zhang, Y. et al, 2022]、放疗[Succony et al, 2021]、化疗[Frost et al, 2020]以及靶向治疗[Kim et al, 2021]和免疫治疗[Song et al, 2022]在内的治疗手段不断完善，肺腺癌患者的生存率有所提高，但总体预后仍然不容乐观，特别是对于晚期肺腺癌患者而言[Jones et al, 2018; Schabath et al, 2019; Zappa et al, 2016]。

在这种背景下，肺腺癌的预后模型研究显得尤为重要。预后模型能够帮助医生更准确地评估患者的生存预期，为患者提供个性化的治疗方案，从而提高疾病的治疗效果和患者的生活质量。近年来，随着生物信息学、人工智能等技术的发展，肺腺癌预后模型的研究取得了显著进展。这些高科技手段的应用，使得预后模型能够更全面地考虑肿瘤的分子生物学特征、患者的临床特征以及治疗响应等多方面因素，从而提高预后评估的准确性[Ali et al, 2023; Wong et al, 2020]。

然而，尽管取得了一定的进展，肺腺癌预后模型的研究仍面临着多重挑战。例如，如何整合和分析来自不同来源的大数据，如何在模型中准确反映肿瘤的异质性，以及如何将预后模型更好地应用于临床实践等，都是当前研究中亟待解决的问题。

相信随着研究的深入和技术的进步，肺腺癌预后模型有望变得更加精准和实用。这不仅能够为患者提供更为个性化的治疗建议，还有助于推动肺腺癌治疗策略的优

化和发展。因此,加强肺腺癌预后模型的研究,对于提高肺腺癌患者的生存率和生活质量具有重要的意义。

因此,在本综述中,我们总结了目前医学研究中主流的构建肺腺癌预后模型的方法,主要从免疫相关因素、各种细胞死亡相关的热点表型、人工智能以及其他因素方面进行阐述,并对每种方法目前的最新进展进行了归纳,以便对目前肺腺癌预后模型的构建有一个整体的理解。

一, 免疫相关因素在肺腺癌预后模型构建中的作用

在构建肺腺癌预后模型中,免疫相关因素扮演着至关重要的角色。这些因素包括但不限于免疫细胞浸润、免疫检查点表达以及细胞因子的水平,它们在肿瘤微环境中的相互作用对于肺腺癌的发展、转移以及患者的生存预后具有重要影响。通过综合分析这些免疫标志物,研究人员可以更准确地预测疾病进展和治疗效果,从而为患者提供个性化的治疗方案。当前的研究表明,将免疫相关因素纳入预后模型可以显著提高模型的预测能力,指导临床治疗决策,优化患者管理策略,从而提高肺腺癌患者的生存率和生活质量。因此,免疫相关因素在肺腺癌预后模型构建中的应用不仅反映了肿瘤生物学的复杂性,也展现了精准医疗的应用前景。

T 细胞存在于肿瘤微环境中,在肿瘤发生发展的各个阶段均发挥着重要作用。研究者们根据肺腺癌的 scRNA-seq 数据并结合 TCGA 数据库构建了由 9 个 T 细胞相关的预后基因组成的风险模型,并证明该预后模型在肺腺癌患者中是独立的预后危险因素[Zhang, J. et al, 2023]。

肿瘤免疫微环境与肿瘤的发生、发展和转移有关。Sun 等研究了肺腺癌免疫相关特征并开发了一个名为 IPS 的四基因标记(ARNTL2、ECT2、PPIA 和 TUBA4A),并证明 IPS 在两个独立的队列中表现优于其他模型[Sun et al, 2020]。Shan 等根据肿瘤微环境相关基因确定了两种亚型。最终选择了 14 个基因用于构建肺腺癌的预后模型,并且建立的模型能够比以前已知的模型更准确地预测肺腺癌患者的预后,为肺腺癌患者预后的预测提供了进一步的见解,并可能指导个体化治疗[Shan et al, 2022]。

免疫治疗在肺腺癌的治疗中发挥着重要的作用。Zhang 等构建了六基因肺腺癌预后模型(PLK1、HMMR、ANLN、SLC2A1、SFTPB 和 CYP4B1)。结果表明肺腺癌患者高风险组和低风险组在生存率、免疫细胞浸润、肿瘤突变负荷、免疫检查点和免疫微环境方面存在显著差异。并且在 GEO 数据集中验证了预后模型的准确性[Zhang, P. et al, 2022]。

二, 细胞死亡相关热点表型在肺腺癌预后模型构建中的作用

在肺腺癌预后模型的构建中,细胞死亡相关的热点表型,如细胞凋亡、坏死、自噬以及程序性细胞死亡(如铁死亡和焦亡)等起着重要作用。这些细胞死亡方式

的调节在肺腺癌的发展、治疗反应以及疾病预后中扮演着关键角色。细胞死亡机制的变化不仅反映了肿瘤的生物行为，也是治疗抵抗发展的重要因素之一。通过深入分析肿瘤样本中的细胞死亡相关标志物，研究人员能够识别出与肺腺癌患者生存率密切相关的靶点，进而提高预后模型的预测准确性。这种方法有助于实现对肺腺癌患者的个性化治疗策略，通过精准预测疾病进展和治疗效果，优化治疗方案，从而提升患者的生存机会和生活质量。

铜中毒可触发细胞死亡机制，包括凋亡和自噬，影响细胞生存。研究者们通过分析研究铜中毒相关基因的预后价值并建立铜中毒相关预后模型。结果显示基于铜中毒特征评分的高危组患者的总生存期均短于低风险组患者。并通过体外实验证明敲低 *BARX1* 和 *GFRA3* 可以增加肺腺癌细胞对铜中毒的敏感性[Chen et al, 2022]。Wang 等研究了铜中毒相关 lncRNA 在临床预后预测和免疫治疗中的作用，以及与药物敏感性的关系，结果显示 16 个铜中毒相关 lncRNA 建立的模型可以准确预测肺腺癌患者的预后，并可能为临床应用和免疫治疗提供新的见解[Wang et al, 2022]。Zhang 等基于铜中毒和免疫相关基因构建了肺腺癌患者生存预测模型，结果显示，一共鉴定了 5 个致癌驱动基因，即 *CD79B*、*PEBP1*、*PTK2B*、*STXBP1* 和 *ZNF671*，并建立了比例风险回归模型。在高危人群中，训练集和验证集的存活率都显著降低。此外，与低风险组相比，高风险组的免疫细胞浸润和免疫检查点表达水平较低[Zhang, W. et al, 2023]。

二硫化物变性是一种新的细胞死亡方式，由积累的细胞内二硫化物与肌动蛋白细胞骨架蛋白结合引起。研究者们探讨了肺腺癌中二硫化物变性相关基因的表达及其预后价值。结果显示，5 个二硫化物变性相关基因（*ACTB*、*FLNB*、*NCKAP1*、*SLC3A2*、*SLC7A11*）纳入最终的模型构建，并且模型可以稳定可靠地预测肺腺癌的预后[Ni et al, 2023]。Huang 等的研究检测了 21 个与二硫化物变性相关的基因，并发现了与肺腺癌预后较差相关的基因特征。同时构建预后模型并使用三个独立的数据集进行验证，在 1 年、3 年和 5 年时 AUC 值均大于 0.5[Huang et al, 2023]。

细胞焦亡是通过炎症小体的激活进行调节，并参与肿瘤的发生。然而，细胞焦亡相关 lncRNA 对肺腺癌预后的影响尚未完全阐明。Song 等发现了 84 个具有差异表达的细胞焦亡相关 lncRNA，并基于其中五种 lncRNA（*GSEC*、*FAM83A-S1*、*AL606489.1*、*AL034397.3* 和 *AC010980.2*）构建了一个新的肺腺癌预后预测模型，该模型在预测预后方面表现出良好的性能，并且与免疫细胞浸润有关。同时准确地预测了患者的总体生存率，具有良好的临床应用价值[Song et al, 2021]。

失巢凋亡是程序性细胞死亡的另一种形式，在癌症侵袭和转移中发挥着关键作用，其作用是防止分离的癌症细胞重新附着到其他基质上进行异常增殖。Diao 等对失巢相关基因在肺腺癌中的预后影响进行全面分析。结果显示，模型在预测肺腺癌

患者的总生存期方面显示出令人满意的预测性能[Diao et al, 2023]。

细胞死亡相关基因指示癌症患者的预后。泛凋亡是一种新观察到的细胞死亡形式,研究人员将其与癌症细胞死亡和抗肿瘤免疫联系起来。即便如此,其在肺腺癌中的意义仍有待阐明。Zhang 等研究发现,基于六个泛凋亡相关基因(NLRC4、FADD、TRADD、RIPK1、PSTPIP2 和 MLKL)构建的预后模型可以准确估计肺腺癌患者的预后[Zhang, C. et al, 2023]。

胞葬作用就是指吞噬细胞清除程序性死亡细胞的过程。越来越多的证据强调了胞葬作用在肿瘤进展中的潜在意义。Liu 等基于胞葬作用相关基因构建肺腺癌的预后风险模型。结果显示,研究建立了以 7 个胞葬作用相关基因为中心的风险预后模型,该模型在预测肺腺癌患者的生存率和免疫状况方面具有较高的准确性[Liu et al, 2023]。

三, 人工智能在肺腺癌预后模型构建中的作用

人工智能(Artificial Intelligence, AI)在肺腺癌预后模型构建中的作用日益凸显,为精准医疗提供了新的视角和方法。利用 AI 技术,如机器学习和深度学习,研究人员能够从大量肺腺癌患者的临床数据、基因表达数据、影像数据中提取有价值的信息。AI 模型通过分析这些数据,能够识别出影响肺腺癌患者预后的关键生物标志物和模式,甚至发现之前未被注意到的潜在生物学机制。这种方法不仅提高了预后模型的准确性和预测能力,还能帮助医生制定更加个性化的治疗方案。此外,AI 技术在处理复杂数据时的高效率和准确性,使其在早期诊断、疾病分型以及治疗效果评估中展现出巨大潜力,为肺腺癌患者的治疗和管理提供了新的策略和工具。

He 等利用机器学习筛选出 ADM2、CDH17、DKK1、PTX3 和 AC145343.1 五个基因作为肺腺癌的潜在免疫相关基因。其中 ADM2 和 AC145343.1 在肺腺癌患者中预后良好。最终提出了一个免疫预后模型来预测肺腺癌患者的总生存率,该模型为肺腺癌患者的免疫治疗提供了新的标志物和思路[He et al, 2023]。Zhang 等通过分析单细胞 RNA 测序数据和大量 RNA 测序数据,基于总共 26 种机器学习算法开发了肿瘤浸润免疫细胞相关特征,并基于该特征评分可以预测五个独立数据集中肺腺癌患者的生存结果[Zhang, N. et al, 2023]。He 等设计了一个个性化的生存预测系统,该系统可以使用三种不同的人工智能算法提供个体生存曲线。该人工智能预测系统可以通过比较不同治疗下的个体死亡率风险曲线,直接传达治疗效果[He, T. et al, 2022]。

Zhang 等利用影像组学特征预测表皮生长因子受体酪氨酸激酶抑制剂治疗后肺腺癌患者无进展生存期。结果显示,构建的影像组学模型显示了肺腺癌疾病进展的最佳预后表现。这种动态成像生物标志物有望监测治疗反应并实现个性化管理[Zhang, X. et al, 2023]。Feng 等基于 H&E 染色图像构建的拓扑细胞图,通过计算空

间相邻肿瘤淋巴细胞数量与肿瘤细胞数量的比值,提出了一种人工智能量化的肿瘤-淋巴细胞空间相互作用评分,通过将该评分与临床病理风险因素相结合建立的模型改进了三个独立队列中无病生存期的预测[Feng et al, 2023]。Zhang 等分析了肺腺癌的程序性细胞死亡相关基因的多组变异。使用 10 种机器学习算法(101 种组合)的交叉验证来综合开发和验证基于程序性细胞死亡相关基因的表达谱的最佳预后细胞死亡评分模型。结果显示,该模型可以准确预测患者的预后及其对治疗的反应[Zhang, Y. et al, 2023]。

四, 其他病理生理因素在肺腺癌预后模型构建中的作用

昼夜节律紊乱是肿瘤增殖的主要因素。但到目前为止,昼夜节律紊乱在肺腺癌的临床预测中的作用尚未明确。He 等研究表明,昼夜节律紊乱在 T 细胞耗竭中起着关键作用,这可能是肺腺癌预后不佳的原因[He, L. et al, 2022]。

广泛的研究已经确定了癌症相关成纤维细胞与癌症发展的各个阶段之间的显著相关性,包括起始、血管生成、进展和对治疗的抵抗。在 Ren 等的研究中,有 3 个癌症相关成纤维细胞簇与肺腺癌的预后显著相关,结合风险特征和临床病理特征的预后模型显示出良好的临床适用性,同时通过体外实验证实了相关基因 EXP1 在促进肺腺癌中肿瘤细胞的侵袭和生长中的作用[Ren et al, 2023]。

RNA 修饰在癌症的发展和进展中具有病理和预后意义,其中 m6A 和 m5C 是具有代表性的调节因子。这些 RNA 修饰可以通过调节基因表达对其他 RNA 的功能产生影响。Tian 等构建了一个由 METTL3、NPLOC4、RBM15、YTHDF1、IGF2BP1、NSUN3 和 NSUN7 组成的七基因 m6A/m5C 风险模型,以对早期肺腺癌的预后进行分层($p=0.0049$, $AUC=0.791$)。证实七基因风险模型可能是早期肺腺癌预后评估的关键工具[Tian et al, 2023]。

肿瘤细胞是否通过基底膜侵袭是决定肺腺癌预后的关键因素。Zhang 等研究共确定了 31 个肺腺癌的预后基底膜相关基因,并成功建立了一个名为 BMscore 的新的 17mRNA 预后模型来预测肺腺癌患者的总生存率。结果显示高 BMscore 组肺腺癌患者预后较差[Zhang, Z. et al, 2023]。

越来越多的证据表明, lncRNA 可以作为癌症的潜在预后因素。Gong 等利用血管生成相关的 lncRNA 作为潜在的预后因素,建立肺腺癌的预后模型。结果显示,研究中鉴定的四种血管生成相关 lncRNA 可作为肺腺癌患者的预后生物标志物[Gong et al, 2023]。

配体和受体在细胞通讯中起着重要作用。Li 等证实了配体-受体对在肺腺癌发展中的重要性。结果表明配体-受体对在肺腺癌的发展中起着关键作用。该预后模型可以预测肺腺癌患者的预后并指导个性化治疗[Li et al, 2022]。

五, 展望

对肺腺癌预后模型的研究在不断进步,展现出前所未有的深度与广度。当前的研究重点在于整合多源数据,包括基因组、转录组、蛋白质组以及临床参数,以构建更加精准和全面的预后评估模型。特别是,随着高通量测序技术的发展和人工智能算法的应用,研究人员能够在更细致的层面上理解肿瘤的生物学特性,识别出新的生物标志物,为肺腺癌的分型和预后提供了新的视角。

此外,影像组学的兴起为肺腺癌的预后模型增添了新的维度。通过分析来自CT、MRI等影像的数据,研究者可以提取出肿瘤的形态学特征,这些特征与肿瘤的基因表达及患者的生存结果密切相关,从而提高预后模型的预测精度。

未来的研究有望通过进一步的跨学科合作,如生物信息学、计算机科学与临床医学的深度融合,推动个性化医疗的实现。特别是,在处理和解析大规模生物医学数据方面,人工智能技术的优化和创新将是关键。同时,随着精准医疗理念的深入人心,预后模型的研究也将更加注重多样性和包容性,以适应不同人群的特定需求。

综上所述,肺腺癌预后模型的研究正处于一个快速发展的阶段,未来有望通过更高效的数据整合、更先进的分析方法和更深入的生物学理解,为肺腺癌患者提供更为精确的预后信息和更优化的治疗策略。

参考文献:

- Ali H et al. (2023). Improving diagnosis and prognosis of lung cancer using vision transformers: a scoping review. *BMC Med Imaging*. 23(1):129.
- Chen Y et al. (2022). Identification and validation of a novel cuproptosis-related signature as a prognostic model for lung adenocarcinoma. *Front Endocrinol (Lausanne)*. 13:963220.
- Diao X et al. (2023). Identification of a novel anoikis-related gene signature to predict prognosis and tumor microenvironment in lung adenocarcinoma. *Thorac Cancer*. 14(3):320-330.
- Feng Z et al. (2023). Artificial intelligence-quantified tumour-lymphocyte spatial interaction predicts disease-free survival in resected lung adenocarcinoma: A graph-based, multicentre study. *Comput Methods Programs Biomed*. 238:107617.
- Frost N et al. (2020). Pemetrexed-Based Chemotherapy Is Inferior to Pemetrexed-Free Regimens in Thyroid Transcription Factor 1 (TTF-1)-Negative, EGFR/ALK-Negative Lung Adenocarcinoma: A Propensity Score Matched Pairs Analysis. *Clin Lung Cancer*. 21(6):e607-e621.
- Gong Q et al. (2023). Construction and validation of an angiogenesis-related lncRNA prognostic model in lung adenocarcinoma. *Front Genet*. 14:1083593.
- He L et al. (2022). Single-cell transcriptomic analysis reveals circadian rhythm disruption associated with poor prognosis and drug-resistance in lung adenocarcinoma. *J Pineal Res*. 73(1):e12803.
- He T et al. (2022). Artificial intelligence predictive system of individual survival rate for lung adenocarcinoma. *Comput Struct Biotechnol J*. 20:2352-2359.
- He X et al. (2023). Machine learning-based immune prognostic model and ceRNA network construction for lung adenocarcinoma. *J Cancer Res Clin Oncol*. 149(10):7379-7392.
- Huang J et al. (2023). Identification of a disulfidptosis-related genes signature for prognostic

- implication in lung adenocarcinoma. *Comput Biol Med.* 165:107402.
- Jones G S et al. (2018). Recent advances in the management of lung cancer. *Clin Med (Lond).* 18(Suppl 2):s41-s46.
- Kim S Y et al. (2021). Modeling Clinical Responses to Targeted Therapies by Patient-Derived Organoids of Advanced Lung Adenocarcinoma. *Clin Cancer Res.* 27(15):4397-4409.
- Li D et al. (2022). Molecular subtyping for lung adenocarcinoma and a novel prognostic model based on ligand-receptor pairs. *Adv Med Sci.* 67(2):316-327.
- Liu J et al. (2023). Construction and validation of an efferocytosis-related prognostic signature in lung adenocarcinoma. *J Cancer Res Clin Oncol.* 149(16):14577-14596.
- Ni L et al. (2023). The expression and prognostic value of disulfidptosis progress in lung adenocarcinoma. *Aging (Albany NY).* 15(15):7741-7759.
- Ren Q et al. (2023). A novel signature predicts prognosis and immunotherapy in lung adenocarcinoma based on cancer-associated fibroblasts. *Front Immunol.* 14:1201573.
- Schabath M B et al. (2019). Cancer Progress and Priorities: Lung Cancer. *Cancer Epidemiol Biomarkers Prev.* 28(10):1563-1579.
- Shan Q et al. (2022). Clustering analysis and prognostic signature of lung adenocarcinoma based on the tumor microenvironment. *Sci Rep.* 12(1):12059.
- Siegel R L et al. (2020). Cancer statistics, 2020. *CA Cancer J Clin.* 70(1):7-30.
- Siegel R L et al. (2021). Cancer Statistics, 2021. *CA Cancer J Clin.* 71(1):7-33.
- Skoulidis F et al. (2019). Co-occurring genomic alterations in non-small-cell lung cancer biology and therapy. *Nat Rev Cancer.* 19(9):495-509.
- Song J et al. (2021). A novel pyroptosis-related lncRNA signature for prognostic prediction in patients with lung adenocarcinoma. *Bioengineered.* 12(1):5932-5949.
- Song P et al. (2022). Identification and Validation of a Novel Signature Based on NK Cell Marker Genes to Predict Prognosis and Immunotherapy Response in Lung Adenocarcinoma by Integrated Analysis of Single-Cell and Bulk RNA-Sequencing. *Front Immunol.* 13:850745.
- Succony L et al. (2021). Adenocarcinoma spectrum lesions of the lung: Detection, pathology and treatment strategies. *Cancer Treat Rev.* 99:102237.
- Sun S et al. (2020). Development and validation of an immune-related prognostic signature in lung adenocarcinoma. *Cancer Med.* 9(16):5960-5975.
- Thai A A et al. (2021). Lung cancer. *Lancet.* 398(10299):535-554.
- Tian L et al. (2023). Prognostic Value and Genome Signature of m6A/m5C Regulated Genes in Early-Stage Lung Adenocarcinoma. *Int J Mol Sci.* 24(7).
- Wadowska K et al. (2020). Genetic Markers in Lung Cancer Diagnosis: A Review. *Int J Mol Sci.* 21(13).
- Wang F et al. (2022). Cuproptosis-related lncRNA predict prognosis and immune response of lung adenocarcinoma. *World J Surg Oncol.* 20(1):275.
- Wong C W et al. (2020). Radiogenomics of lung cancer. *J Thorac Dis.* 12(9):5104-5109.
- Zappa C et al. (2016). Non-small cell lung cancer: current treatment and future advances. *Transl Lung Cancer Res.* 5(3):288-300.
- Zhang C et al. (2023). Identifying prognostic genes related PANoptosis in lung adenocarcinoma and developing prediction model based on bioinformatics analysis. *Sci Rep.* 13(1):17956.
- Zhang J et al. (2023). T cell-related prognostic risk model and tumor immune environment modulation in lung adenocarcinoma based on single-cell and bulk RNA sequencing. *Comput*

- Biol Med. 152:106460.
- Zhang N et al. (2023). An artificial intelligence network-guided signature for predicting outcome and immunotherapy response in lung adenocarcinoma patients based on 26 machine learning algorithms. *Cell Prolif.* 56(4):e13409.
- Zhang P et al. (2022). Analysis of prognostic model based on immunotherapy related genes in lung adenocarcinoma. *Sci Rep.* 12(1):22077.
- Zhang W et al. (2023). Identification of cuproptosis and immune-related gene prognostic signature in lung adenocarcinoma. *Front Immunol.* 14:1179742.
- Zhang X et al. (2023). Prognostic analysis and risk stratification of lung adenocarcinoma undergoing EGFR-TKI therapy with time-serial CT-based radiomics signature. *Eur Radiol.* 33(2):825-835.
- Zhang Y et al. (2022). Surgery for pre- and minimally invasive lung adenocarcinoma. *J Thorac Cardiovasc Surg.* 163(2):456-464.
- Zhang Y et al. (2023). A programmed cell death-related model based on machine learning for predicting prognosis and immunotherapy responses in patients with lung adenocarcinoma. *Front Immunol.* 14:1183230.
- Zhang Z et al. (2023). A novel basement membrane-related gene signature for prognosis of lung adenocarcinomas. *Comput Biol Med.* 154:106597.

基金资助

国家重点研发计划：人类蛋白质组计划 2.0-蛋白质组学驱动精准医学计划

(Grant Nos. 2020YFE02022200)