

• 慢病专题:癌症 •

肺腺癌中 COL1A1 表达及其功能的生物信息学分析*

邢 龙¹,吴双丽²,袁胜利¹,李保健¹,徐敬宣¹,王丽娟¹,喻 光^{1△}
(青岛滨海学院附属医院:1.肿瘤防治中心;2.神经电生理科,山东 青岛 266500)

[摘要] **目的** 通过生物信息学技术探讨 I 型胶原蛋白 $\alpha 1$ 链(COL1A1)在肺腺癌组织中的表达特征及其与患者预后之间的关联,以及其潜在的生物学功能。**方法** 利用 GEPIA 数据库分析肺腺癌和正常肺组织 COL1A1 表达差异并通过 HPA 数据库进行验证。构建 ROC 曲线分析 COL1A1 在肺腺癌预后中的预测价值。绘制 Kaplan-Meier 曲线以进行生存分析,然后通过 String 数据库得出与 COL1A1 关系最密切的前 10 个蛋白(COL1A2、COL3A1、COL5A2、SPARC、LUM、PCOLCE、ADAMTS2、BMP1、ITGB1、RUNX2)。通过 R 软件分析相关蛋白的生物学功能。**结果** 相比于正常肺组织,COL1A1 在肺腺癌组织中显著高表达($P<0.01$)。ROC 曲线显示,COL1A1 在肺腺癌和正常肺组织之间有良好的预测价值。与 COL1A1 低表达组相比,COL1A1 高表达组总体生存期较差($P<0.05$)。与 COL1A 关系密切的前 10 个蛋白主要在胶原纤维组织、原纤维胶原三聚体、细胞外基质结构成分、蛋白质消化和吸收等方面发挥作用。**结论** COL1A1 的异常表达参与了肺腺癌发生、发展过程,或可成为肺腺癌的诊断、治疗靶点,但其在肺腺癌中发挥的具体功能尚需进一步体外研究。

[关键词] 肺腺癌; I 型胶原蛋白 $\alpha 1$; 生物信息学; 预后; 数据库
DOI:10.3969/j.issn.1009-5519.2025.11.004 **中图法分类号:**R734.2
文章编号:1009-5519(2025)11-2519-05 **文献标识码:**A

Bioinformatics analysis of COL1A1 expression and function in lung adenocarcinoma*
XING Long¹,WU Shuangli²,YUAN Shengli¹,LI Baojian¹,XU Jingxuan¹,WANG Lijuan¹,YU Guang^{1△}
(1. Department of Oncology;2. Department of Neuroelectrophysiology, Affiliated Hospital of Qingdao Binhai University Hospital, Qingdao, Shandong 266500, China)

[Abstract] **Objective** To analyze the expression of COL1A1 in lung adenocarcinoma tissues and its relationship with prognosis and biological function through bioinformatics techniques. **Methods** The expression differences of COL1A1 in lung adenocarcinoma and normal lung tissues were analyzed using the GEPIA database and verified through the HPA database. The ROC curve was constructed to analyze the predictive value of COL1A1 in the prognosis of lung adenocarcinoma. Draw the Kaplan-Meier curve for survival analysis. Then,through the String database,the top 10 proteins most closely related to the COL1A1 protein(COL1A2, COL3A1, COL5A2, SPARC, LUM, PCOLCE, ADAMTS2, BMP1, ITGB1, RUNX2) were obtained. The biological functions of the relevant proteins were analyzed through R software. **Results** Compared with normal lung tissues,COL1A1 was significantly highly expressed in lung adenocarcinoma tissues($P<0.01$). The ROC curve showed that COL1A1 had a good predictive value between lung adenocarcinoma and normal lung tissue. Compared with the low-expression group of COL1A1,the OS of the high-expression group of COL1A1 was poorer($P<0.05$). The top 10 proteins closely related to COL1A mainly play a role in collagen fiber tissue, fibrillary collagen trimer, extracellular matrix structural components, protein digestion and absorption, etc. **Conclusion** The abnormal expression of COL1A1 is involved in the occurrence and development of lung adenocarcinoma, and may be a target for diagnosis and treatment of lung adenocarcinoma. However, the specific function of COL1A1 in lung adenocarcinoma needs to be further studied in vitro.

[Key words] Lung adenocarcinoma; COL1A1; Bioinformatics; Prognosis; Database

* 基金项目:青岛滨海学院附属医院临床科学研究专项项目(2024KYZ03;2024KYY06)。
作者简介:邢龙(1986—),硕士研究生,主治医师,主要从事恶性肿瘤的基础与临床研究。△ 通信作者,E-mail: yuguang8848@163.com。
网络首发 [https://link.cnki.net/urlid/50.1129.R.20250619.1547.027\(2025-06-19\)](https://link.cnki.net/urlid/50.1129.R.20250619.1547.027(2025-06-19))

根据我国国家癌症中心的最新数据,我国新发肺癌患者约为 106.06 万例/年,其中男 65.87 万例/年,女 40.19 万例/年,因肺癌死亡患者约为 73.33 万例/年,其中男 51.59 万例/年,女 21.74 万例/年^[1]。另外,肺癌的主要病理组织学亚型的占比亦在改变,大约从 2000 年以后,肺癌患者最常见的病理组织学亚型已从以前的鳞状细胞癌转变为腺癌。随着肺腺癌发病机制研究的逐渐深入,新治疗方法取得了较大的突破,但中晚期肺腺癌的 5 年生存率依然不容乐观^[2]。因此,针对其发病机制的分子层面的研究始终未曾中断。

I 型胶原蛋白 α1 链(COL1A1)存在于大多数结缔组织中。研究发现,COL1A1 在乳腺癌、非小细胞肺癌、卵巢癌、宫颈癌等组织中有异常表达,且与患者的生存、预后相关,但对肺腺癌组织中 COL1A1 的表达、功能、生物学意义研究较少。本研究通过生物信息学分析 COL1A1 在肺腺癌中表达及功能,揭示 COL1A1 在肺腺癌发生、发展中扮演的角色,旨在为肺腺癌的发病机制、临床治疗及靶向药物的研发提供理论支持。

1 材料与方法

1.1 COL1A1 在肺腺癌中表达水平 GEPIA (<http://gepia.cancer-pku.cn>) 是将 TCGA 数据进行可视化的一款在线分析工具。通过在数据库基因名称检索栏中输入 COL1A1,数据集类别选择为“LU-AD”,其余条件设置设为网站默认,可获得 COL1A1 在肺腺癌及癌旁组织中表达水平。

1.2 COL1A1 蛋白在肺腺癌和正常肺组织中的表达验证 HPA 数据库 (<https://www.proteinatlas.org>) 是一个使用各种组学技术的集成来绘制人类细胞、组织和器官中蛋白质的在线数据库。通过 HPA 数据库可验证 COL1A1 蛋白在肺腺癌和正常肺组织中的表达差异。

1.3 基于 TCGA 数据构建受试者工作特征(ROC)曲线 ROC 曲线可用于评价、比较诊断性试验的效果是否有应用价值。因此,本文通过 ROC 曲线评估以 COL1A1 表达水平区分肺腺癌组织与正常肺组织的可行性。使用 R (3.6.3 版本) 中 pROC 包 (1.17.0.1 版本) 进行分析,并用 ggplot2 包 (3.3.3 版本) 进行可视化,分析 COL1A1 区分肺腺癌和正常肺组织的价值。

1.4 生存分析 根据 COL1A1 表达水平的中位数,将 Kaplan-Meier Plotter(<http://kmplot.com/analysis>) 数据库中的肺腺癌患者分为高、低表达组,评估 2 组患者肺腺癌样本中 COL1A1 表达水平和总体生存期(OS)之间的关系,以 $P < 0.05$ 为差异有统计学意义。

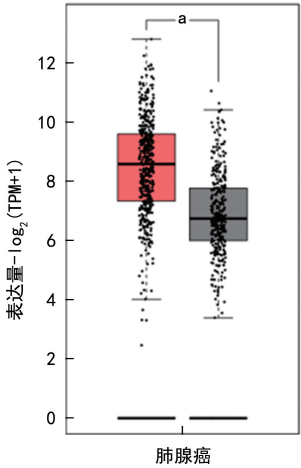
1.5 COL1A1 蛋白相互作用网络及功能分析 通过 String 数据库 (<https://cn.string-db.org/>) 可分析

COL1A1 蛋白的上下游调控蛋白及蛋白间的相关调控关系,选取与 COL1A1 蛋白关系最密切的前 10 个蛋白,构建网络图,并对它们与 COL1A1 蛋白的相互作用关系进行深入分析。

1.6 相关蛋白功能富集分析 使用 R (3.6.3 版本) 中 clusterProfiler 包 (3.14.3 版本) 对与 COL1A1 蛋白关系最密切的前 10 个蛋白进行分析,并用 ggplot2 包 (3.3.3 版本) 进行可视化分析,进行基因本体(GO)功能注释和 KEGG(Kyoto Encyclopedia of Genes and Genomes)通路富集分析。

2 结果

2.1 COL1A1 在肺腺癌中表达 通过对 GEPIA 数据库中 483 例肺腺癌组织标本及 347 例正常肺组织标本分析发现,肺腺癌组织中 COL1A1 蛋白表达水平显著高于正常肺组织($P < 0.01$),见图 1。



注:^a $P < 0.05$ 。
图 1 COL1A1 在肺腺癌及正常肺组织中的表达

2.2 肺腺癌中 COL1A1 蛋白在 HPA 数据库中的表达 HPA 数据库中肺腺癌与正常肺组织人 COL1A1 蛋白的免疫组化结果(图 2)。正常肺组织中 COL1A1 蛋白表达水平检测条件为抗体 HPA011795;染色:低;强度:中度;数量 $< 25\%$;肺腺癌组织中 COL1A1 蛋白表达水平检测条件为抗体 HPA011795;染色:中;强度:中等;数量 $> 75\%$ 。

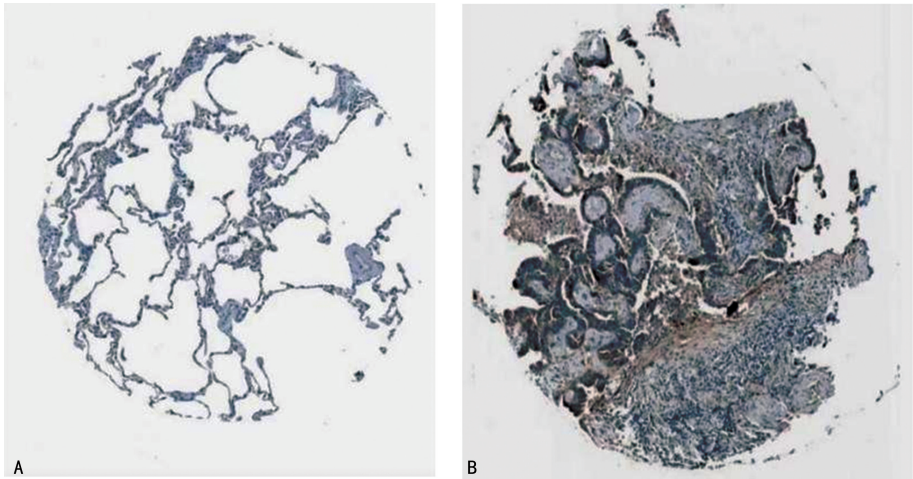
2.3 基于 TCGA 数据 COL1A1 绘制 ROC 曲线 通过绘制 ROC 来评价 COL1A 表达水平并区分肺腺癌组织与正常肺组织的可行性。使用 R 软件 pROC 包 (1.17.0.1 版本) 进行分析,并用 ggplot2 包 (3.3.3 版本) 进行可视化,分析 COL1A1 在肺癌预后预测中的价值,得出 ROC 曲线下面积(AUC)为 0.898,表示在预测肺腺癌组织和正常肺组织结局上,变量 COL1A1 的预测能力有较好的准确性(图 3)。

2.4 生存分析 如图 4 所示,和 COL1A1 低表达组相比,COL1A1 高表达组的 OS 明显变差,差异有统计学意义($P < 0.05$)。

2.5 COL1A1 蛋白相互作用网络及功能分析 通过

String 线数据库,探索与 COL1A1 蛋白关系最密切的前 10 个蛋白,结果表明,COL1A1 蛋白与 COL1A2、

COL3A1、COL5A2、SPARC、LUM、PCOLCE、AD-AMTS2、BMP1、ITGB1、RUNX2 关系密切(图 5)。



注:A. COL1A1 在正常肺组织中的蛋白表达水平;B. COL1A1 在肺腺癌组织中的蛋白表达水平。

图 2 COL1A1 蛋白在 HPA 数据库中的表达

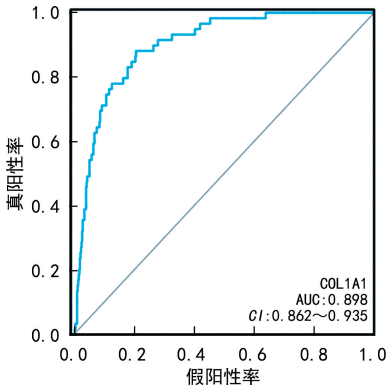


图 3 肺腺癌中 COL1A1 的 ROC 曲线

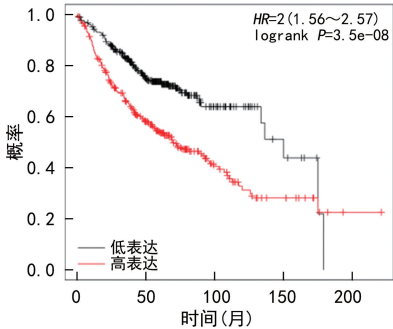


图 4 COL1A1 与肺腺癌患者的生存关系

2.6 COL1A1 及相关蛋白的富集分析 GO 功能富集分析表明,生物过程(BP)在胶原纤维组织、细胞外基质(ECM)及细胞对氨基酸刺激的反应方面表现出显著的富集;在细胞组成(CC)层面,主要富集于原纤维胶原三聚体、带状胶原纤维、胶原蛋白三聚体复合物及其他细胞外基质成分;而分子功能(MF)则主要集中在细胞外基质结构成分、赋予抗拉强度的细胞外基质成分、血小板衍生生长因子结合及蛋白酶结合等方面(图 6A)。在 KEGG 通路分析中,主要富集于蛋白质的消化与吸收、血小板活化、蛋白聚糖中的生物

ECM-受体相互作用,以及糖尿病并发症中的 AGE-RAGE 信号通路(图 6B)。

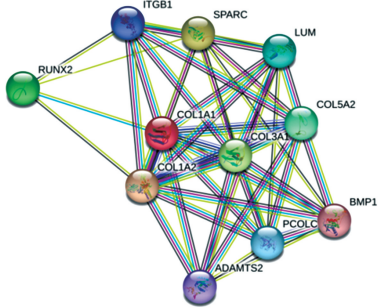


图 5 COL1A1 及相关蛋白 String 蛋白网络

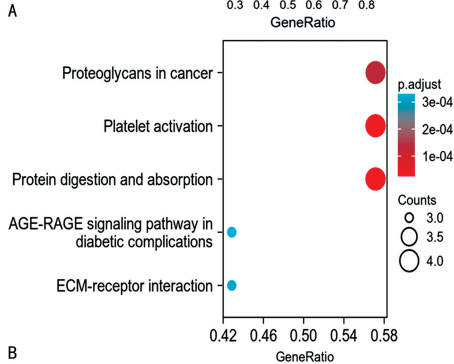
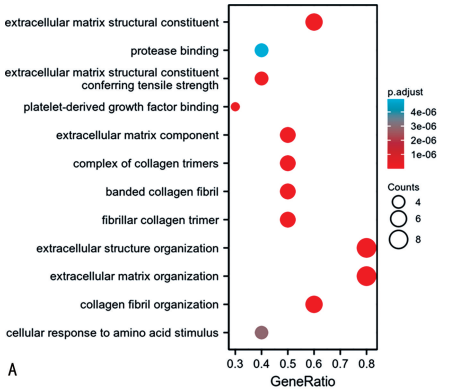


图 6 相关基因 GO(A)和 KEGG(B)富集分析

3 讨 论

肺腺癌是迄今为止最常见的非小细胞肺癌亚型，主要出现在更远端的气道，发病数量约占非小细胞肺癌的 50%^[3-4]。由于其诊断较晚且异质性高，是世界上最致命的癌症之一。

在我国，每年新确诊的肺癌病例超过 70 万例，死亡人数则超过 60 万例。与其他类型的肺癌相比，肺腺癌在非吸烟人群中的发病率显著增加，其发病机制也较为复杂。对于肺腺癌相关生物标志物的研究，如 EGFR、ALK、ROS1、RET、MET、NTRK、KRAS、BRAF、HER2、TP53、KEAP1 和 STK11 等^[5-9]，为改善患者预后提供了重要依据。随着现有治疗水平的提升，肺癌患者的生存期得到了显著延长，然而，原位期、局部期及晚期肺癌患者的 5 年生存率依然仅为 60%、33%和 6%。

本文通过利用公共数据库对 COL1A1 的生物信息学分析发现，COL1A1 在肺腺癌组织中显著高表达，差异有统计学意义($P<0.05$)。ROC 曲线结果提示变量 COL1A1 的预测能力有较好的准确性，且与肺腺癌患者中 COL1A1 低表达者相比，高表达者生存预后显著变差($P<0.05$)。进一步分析发现，与 COL1A1 蛋白相关度最高的前 10 个蛋白功能主要在胶原纤维组织、胶原纤维组织细胞外基质结构成分、蛋白质消化和吸收等方面富集。提示 COL1A1 可能在肺腺癌的发生、发展中扮演着重要角色。

研究发现，COL1A1 的过度表达可增加肿瘤组织的刚性及机械应力，但增加的机械应力会导致血管受压，导致肿瘤组织的低灌注和药物输送受损^[10]。COL1A1 是胶质母细胞瘤间充质转化动态组织的核心基因，也是胶质母细胞瘤恶性行为的强大调节因子，抑制 COL1A1 可以消除癌瘤，重新编程恶性组织病理学表型，减少间充质相关基因的表达，诱导肿瘤微环境的变化并延长生存^[11]。COL1A1 的表达水平与免疫相关细胞的浸润水平正相关，提示 COL1A1 可能是肺癌患者潜在的免疫相关生物标志物和治疗靶点^[12]。LIU 等^[13]的研究表明，COL1A1 基因的表达下调能够抑制乳腺癌细胞的转移现象。同时，COL1A1 的细胞表达与雌激素受体(ER)和孕激素受体(PR)的表达水平及转移状态存在一定关联。更重要的是，COL1A1 的高水平表达与患者的生存率降低呈正相关，尤其是在 ER 阳性的乳腺癌患者中。然而，尽管 COL1A1 表达水平较高的患者生存率较低，他们在顺铂化疗中的反应却表现得更加良好。另有研究发现，COL1A1 在原发性肝癌(HCC)肿瘤组织中显著上调，通过沉默 COL1A1 表达可显著抑制了 HCC 细胞的克隆原性、运动性、侵袭性^[14]。相较于 COL1A1

的低表达的胃癌患者，高表达胃癌患者的 OS 较低^[15]。Fostamatinib 对 COL1A1 基因有抑制作用，因此有可能成为胃癌的潜在治疗药物^[16]。

此外，亦有研究发现，COL1A1 与卵巢癌细胞^[17]及非小细胞肺癌细胞^[18]的化疗耐药、宫颈癌^[19]和乳腺癌^[20]细胞的放疗抵抗有关。

ECM 作为细胞与组织之间的生物支持框架，能够引发一系列生物化学和物理信号，进而对细胞的附着、迁移、组织形态发生、修复和血管生成等过程产生重要影响^[21]。已有研究表明，胶原蛋白作为 ECM 的主要组成部分，不管沉积增加、减少都与肿瘤的恶性程度增加相关，因此，COL1A1 可能在肺腺癌发生、发展中发挥重要作用，但其在肺腺癌发生、发展中的作用尚需体外实验进行验证，具体的分子机制亦需进一步研究。

尽管本研究已取得了一定的成果，然而亦充分认识到其局限性。目前的研究主要依赖于数据库的分析，尚未对实际的临床样本进行验证。作者对此限制深感关注，并计划在未来的研究中，专注于收集和分析肺腺癌患者的临床样本，以进一步明确 COL1A1 在肺腺癌中的表达模式及其与患者预后之间的关系。同时，还打算深入探讨 COL1A1 在肺腺癌的发生及进展中的具体作用机制，旨在为肺腺癌的发病机制研究和临床治疗提供更加坚实和全面的理论基础。

参考文献

[1] 郑荣寿,陈茹,韩冲峰,等. 2022 年中国恶性肿瘤流行情况分析[J]. 中华肿瘤杂志, 2024, 46(3): 221-231.

[2] LOPICCOLO J, GUSEV A, CHRISTIANI D C, et al. Lung cancer in patients who have never smoked- an emerging disease[J]. Nat Rev Clin Oncol, 2024, 21(2): 121-146.

[3] MANTILLA J G, MOREIRA A L. The grading system for lung adenocarcinoma: brief review of its prognostic performance and future directions[J]. Adv Anat Pathol, 2024, 31(5): 283-288.

[4] ZHANG J, ZHANG J, YUAN C, et al. Establishment of the prognostic index reflecting tumor immune microenvironment of lung adenocarcinoma based on Metabolism-Related genes[J]. J Cancer, 2020, 11(24): 7101-7115.

[5] CIARDIELLO F, HIRSCH F R, PIRKER R, et al. The role of anti-EGFR therapies in EGFR-TKI-resistant advanced non-small cell lung cancer[J]. Cancer Treat Rev, 2024, 122: 102664.

[6] MAIONE P, PALMA V, PUCILLO G, et al. Targeting ALK receptors in non-small cell lung cancer: what is the road ahead? [J]. Expert Opin Ther Targets, 2024, 28(8): 659-668.

[7] YANG X, TANG Z, LI J, et al. Progress of non-small-cell lung cancer with ROS1 rearrangement[J]. Front Mol Biosci, 2023, 10: 1238093.

[8] MARINELLI D, SIRINGO M, METRO G, et al. Non-small-cell lung cancer: how to manage ALK-, ROS1- and NTRK-rearranged disease[J]. Drugs Context, 2022, 11: 2022-3-1.

[9] OLMEDO M E, CERVERA R, CABEZON G L, et al. New horizons for uncommon mutations in non-small cell lung cancer: BRAF, KRAS, RET, Met, NTRK, HER2 [J]. World J Clin Oncol, 2022, 13(4): 276-286.

[10] MENG C, HE Y P, WEI Z Q, et al. MRTF-A mediates the activation of COL1A1 expression stimulated by multiple signaling pathways in human breast cancer cells[J]. Biomed Pharmacother, 2018, 104: 718-728.

[11] COMBA A, FAISAL S M, DUNN P J, et al. Spatiotemporal analysis of glioma heterogeneity reveals COL1A1 as an actionable target to disrupt tumor progression[J]. Nat Commun, 2022, 13(1): 3606.

[12] GENG Q S, SHEN Z B, LI L F, et al. COL1A1 is a prognostic biomarker and correlated with immune infiltrates in lung cancer[J]. Peer J, 2021, 9: e11145.

[13] LIU J, SHEN J X, WU H T, et al. Collagen 1a1 (COL1A1) promotes metastasis of breast cancer and is a potential therapeutic target[J]. Discov Med, 2018, 25(139): 211-223.

[14] MA H P, CHANG H L, BAMODU O A, et al. Collagen 1a1 (COL1A1) is a reliable biomarker and putative therapeutic target for hepatocellular carcinogenesis and metastasis[J]. Cancers (Basel), 2019, 11(6): 786.

[15] WANG D W, SU F, YANG L J, et al. Bioinformatics analysis and identification of potential genes associated with pathogenesis and prognosis of gastric cancer [J]. Curr Med Sci, 2022, 42(2): 357-372.

[16] SALARIKIA S R, KASHKOOL M, TAGHIPOUR M J, et al. Identification of hub pathways and drug candidates in gastric cancer through systems biology[J]. Sci Rep, 2022, 12(1): 9099.

[17] YANG F, ZHAO Z Y, CAI S Y, et al. Detailed molecular mechanism and potential drugs for COL1A1 in carboplatin-resistant ovarian cancer [J]. Front Oncol, 2020, 10: 576565.

[18] JIA R, WANG C L. MiR-29b-3p reverses cisplatin resistance by targeting COL1A1 in non-small-cell lung cancer a549/DDP cells [J]. Cancer Manag Res, 2020, 12: 2559-2566.

[19] LIU S R, LIAO G W, LI G W. Regulatory effects of COL1A1 on apoptosis induced by radiation in cervical cancer cells [J]. Cancer Cell Int, 2017, 17: 73.

[20] KIM K, KIM Y J. RhoBTB3 regulates proliferation and invasion of breast cancer cells via coll1a1 [J]. Mol Cells, 2022, 45(9): 631-639.

[21] WANG Z, ZHANG F, WANG Z, et al. Hierarchical assembly of bioactive amphiphilic molecule pairs into supramolecular nanofibril Self-Supportive scaffolds for stem cell differentiation [J]. J Am Chem Soc, 2016, 138(45): 15027-15034.

(收稿日期: 2025-02-26 修回日期: 2025-06-29)

(上接第 2518 页)

[9] MU Y, XIE F, SUN T. Clinical value of seven autoantibodies combined detection in the diagnosis of lung cancer [J]. J Clin Lab Anal, 2020, 34(8): e23349.

[10] LI S, MA Y, XIONG Y, et al. Five tumor-associated autoantibodies expression levels in serum predict lung cancer and associate with poor outcome [J]. Translat Cancer Res, 2019, 8(4): 1364-1373.

[11] DU Q, YU R, WANG H, et al. Significance of tumor-associated autoantibodies in the early diagnosis of lung cancer [J]. Clin Respir J, 2018, 12(6): 2020-2028.

[12] ZHANG R, MA L, LI W, et al. Diagnostic value of multiple tumor-associate autoantibodies in lung cancer [J]. Onco Targets Ther, 2019, 12: 457-469.

[13] CHEN P, LU W, CHEN T. Seven tumor-associated autoantibodies as a serum biomarker for primary screening of early-stage non-small cell lung cancer [J]. J Clin Lab Anal, 2021, 35(11): e24020.

[14] 赵娟, 刘宁, 郭婧婧, 等. 7 种肿瘤相关抗原自身抗体检测在肺癌诊断中的意义 [J]. 临床检验杂志, 2017, 35(5): 326-329.

[15] 邹燕萍, 章于潇, 陈海欣, 等. 7 项肿瘤相关自身抗体在早期非小细胞肺癌中的诊断价值 [J]. 浙江医学, 2021, 43(18): 1988-1991, 1995.

[16] 唐星星. 七种自身抗体检测在肺癌诊断与筛查中的价值 [D]. 昆明: 昆明医科大学, 2020.

[17] DAI L, TSAY J C, LI J, et al. Autoantibodies against tumor-associated antigens in the early detection of lung cancer [J]. Lung Cancer, 2016, 99: 172-179.

(收稿日期: 2025-03-16 修回日期: 2025-07-23)