

中文摘要

基于 TCGA 数据库和 LASSO 回归筛选胃癌预后免疫相关的 lncRNA 及预后模型的建立

目的:

本研究使用生信分析,对癌症基因组图谱(TCGA)数据库中的胃癌转录组数据进行分析,识别与预后相关的免疫相关 lncRNAs(长链非编码 RNA),采用 Cox 回归分析和 LASSO 回归算法构建胃癌预后模型并进行验证。

方法:

利用 TCGA 数据库获取胃癌患者的 RNA 数据,并通过人类参考基因组筛选出 lncRNA 表达数据。然后,从免疫数据库提取免疫基因,利用基因共表达网络得到免疫相关的 lncRNAs,并通过差异分析得到差异表达的免疫相关 lncRNAs。随后,采用 Cox 回归和 LASSO 回归算法,得到关键 lncRNAs 并构建预后风险模型。采用风险模型评分中位值将患者分为胃癌高危组和低危组,通过生存分析验证生存差异,并利用内部数据集进行模型的验证。最后,进行危险评分和临床因素的单变量和多变量 Cox 分析,评估风险得分是否能作为胃癌的独立预后因素。

结果:

1.经过单因素 Cox 回归分析方法,获得了 28 个与胃癌预后相关的免疫 lncRNAs。

2.使用多因素 Cox 回归分析方法确定了 5 个 lncRNAs 用于建立风险模型，其中 AL590666.2 被认为是 $HR < 1$ 的保护性效应因子，AP000695.2、CD44-AS1、GLIS3-AS1、AL356417.2 被认为是 $HR > 1$ 的风险效应因子 ($p < 0.05$)。

3.基于风险模型计算的评分将胃癌划分为高低危组，两组之间的患者生存状态的比较具有统计学显著性 ($p < 0.001$)。

4.Cox 回归显示风险得分可作为胃癌患者的独立预后因素。

结论:

通过利用 TCGA 数据库中胃癌患者的数据和免疫基因库，使用生物信息学方法筛选出与胃癌患者预后相关的免疫相关 lncRNAs，并构建了一种胃癌患者风险得分模型。

根据该模型的风险得分公式，可以将胃癌患者分为高风险组和低风险组，并通过该模型帮助预测胃癌患者的预后。

关键词:

胃癌，免疫，lncRNA，预后模型，TCGA

Abstract

Establishment of a prognostic immune-related lncRNA screening and prognostic model for gastric cancer based on TCGA database and LASSO regression

Objective

In this study, data mining was performed using gastric cancer data from The Tumor Genome Atlas (TCGA, The Cancer Genome Atlas) database, and immune-related lncRNAs (long-stranded non-coding RNAs) associated with prognosis were screened using bioinformatics analysis, then, Cox regression analysis and LASSO regression algorithm were used to construct a prognostic model for gastric cancer and validate it.

Methods

RNA data from gastric cancer patients were obtained using the TCGA database, and lncRNA expression data were screened by the human reference genome. The immune genes were taken from an immune database, and the immune-related lncRNAs were acquired through a gene co-expression network. Differential analysis identified immune-related lncRNAs with differential expression. Afterward, a prognostic risk model was constructed by utilizing Cox regression and LASSO regression algorithms to determine key lncRNAs. The risk model score was utilized to categorize gastric cancer patients into high-risk and low-risk groups based on the median value. Subsequently, survival

analysis was conducted to verify the differences in survival, and the model's validity was confirmed using internal data sets. Finally, univariate and multivariate Cox analyses were performed to evaluate the potential of the risk score as an independent prognostic factor for gastric cancer, in conjunction with each clinical factor.

Results

1. 28 immune lncRNAs associated with prognosis of gastric cancer were obtained after one-way Cox regression analysis method.

2. 5 lncRNAs were identified for risk modeling using a multifactorial Cox regression analysis method, in which AL590666.2 was considered a protective effector for $HR < 1$, in which AP000695.2, CD44-AS1, GLIS3-AS1, and AL356417.2 were considered risk effectors for $HR > 1$ ($p < 0.05$).

3. The risk model scores categorized gastric cancer patients into high and low risk groups, and the survival status disparity between the groups was found to be statistically significant ($p < 0.001$).

4. Cox regression showed that the risk score could be an independent prognostic factor for gastric cancer patients.

Conclusion

A risk score model for gastric cancer patients was constructed by screening immune-related lncRNAs associated with patient prognosis. This was accomplished using bioinformatics methods and data from the

TCGA database and immune gene library.

According to the risk score formula of this model, gastric cancer patients can be divided into high-risk and low-risk groups, and the model can help predict the prognosis of gastric cancer patients.

Keywords

Gastric cancer, immunity, lncRNA, prognostic model, TCGA

中英文对照表

英文缩写	英文全名	中文全名
TCGA	The cancer Genome Atlas	癌症基因组图谱
GC	Gastric cancer	胃癌
NcRNA	Non-coding RNA	非编码核糖核酸
LncRNA	Long non-coding RNA	长链非编码核糖核酸
SncRNA	Small non-coding RNA	小非编码核糖核酸
TME	The tumor microenvironment	肿瘤微环境
TAM	Tumor-associated macrophages	肿瘤相关巨噬细胞
HGP	Human Genome Project	人类基因组计划
HR	Hazard Ratio	风险比
ROC	Receiver operating characteristic	受试者工作特征曲线

目 录

第 1 章 引言	1
第 2 章 综述	3
2.1 胃癌	4
2.2 lncRNA	6
2.3 lncRNA 与肿瘤免疫微环境	7
2.4 胃癌中的 lncRNA	9
2.4.1 HOTAIR	9
2.4.2 H19	10
2.5 TCGA 数据库	11
2.6 小结	13
第 3 章 资料与方法	14
3.1 资料来源	14
3.2 获取免疫相关 lncRNAs	14
3.3 筛选差异表达的免疫相关 lncRNAs 并随机分组	14
3.4 获取与胃癌预后相关的免疫相关 lncRNAs	15
3.5 胃癌预后免疫相关 lncRNAs 预后模型的构建	15
3.6 胃癌预后模型评估和验证	15
第 4 章 结果	17
4.1 差异表达的免疫相关 lncRNAs	17
4.2 胃癌预后模型的建立与可视化	18

4.3 胃癌预后风险模型的评估与验证	24
第 5 章 讨论	28
第 6 章 结论	33
参考文献	34
作者简介	47
致 谢	48

第 1 章 引言

胃癌（Gastric cancer, GC）是全球第四大最为普遍的癌症类型，且是第三大致命的癌症，仅次于肺癌和结肠直肠癌^[1, 2]。在 2020 年的全球卫生组织的一项研究中，胃癌在全球范围内每年新增约 100 万病例，值得注意的是，中国的新增病例占全球新增病例的 43.9%，死亡病例占全球死亡病例的 48.6%^[2, 3]。目前胃癌的综合治疗已发展到包括手术切除、化疗、放疗、靶向和免疫治疗多学科规范化和个体化治疗。但由于早期胃癌诊断率低和晚期胃癌化学药物耐药等诸多因素影响，胃癌 5 年生存率仅 20% 左右。目前发展起来的免疫治疗方法是通过重塑肿瘤免疫微环境来达到抑制肿瘤细胞的作用。CTLA-4 和 PD-1 / PD-L1 等阻断免疫检查点的药物在多种癌症的治疗中显示出巨大潜力，尤其是给选择有限的晚期癌症患者带来了希望。然而由于肿瘤微环境具有高度特异性以及药物的耐药性导致部分患者不能从中受益或产生药物抵抗。研究发现，lncRNA 在多种免疫细胞类型中特异性表达，调节免疫细胞的发育、分化和活化，在肿瘤免疫微环境的调节中起重要作用。作为癌症诊疗领域的热点，探索 lncRNA 对癌症精准个体化治疗具有重要意义。基于高通量测序技术和生物信息学的高速发展，关于胃癌预后基因的研究已有所进展。免疫系统和 lncRNA 作为影响胃癌发展的两个重要因素，本文的研究目的是以免疫相关 lncRNAs 构建一个胃癌患者预后的风险得分模型。

本研究使用了 TCGA 数据库提供的胃癌患者转录组数据和临床

预后数据。利用生物信息学方法筛选胃癌患者肿瘤组织中差异表达的免疫相关 lncRNAs 基因。通过 Cox 回归和 LASSO 回归得到免疫相关 lncRNAs，本研究利用模型公式患者进行风险评分计算，并将其分为高风险组和低风险组，接着比较高、低风险组的生存状况，以检测是否存在差异。此外，我们还采用生存分析方法对模型进行评估，并在内部数据集中验证模型的预测性能。本研究将为胃癌个体化免疫治疗临床试验提供一定的研究方向和理论依据。

第2章 综述

胃癌免疫相关的 lncRNA 研究进展

胃癌是全球第四大最为普遍的癌症类型，且是第三大致命的癌症，仅次于肺癌和结肠直肠癌^[1, 2]。在 2020 年的全球卫生组织的一项研究中，胃癌在全球范围内每年新增约 100 万病例，值得注意的是，中国的新增病例占全球新增病例的 43.9%，死亡病例占全球死亡病例的 48.6%^[2, 3]。目前，胃癌的治疗已经发展到了包括手术切除、化疗、靶向治疗和免疫治疗等多学科联合的规范化和个体化治疗阶段。由于早期胃癌诊断率低和晚期胃癌化学药物耐药等多种因素的影响，胃癌 5 年生存率仅约 20%。近年来，随着检测技术的不断更新优化和包含患者基因组数据和临床数据的公共数据库的蓬勃发展，学者们正在通过基因水平研究胃癌肿瘤微环境的特点和机制，以更深入的认识来推进肿瘤治疗的研究。其中，免疫治疗已成为研究热点。PD-1 和 PD-L1 等免疫检查点抑制剂的问世为癌症晚期患者带来了希望。除此之外，lncRNA 作为肿瘤的重要调节因子，不仅可以直接影响肿瘤的进展，还通过肿瘤免疫微环境，间接作用于肿瘤生长和转移。具体而言，lncRNA 可以影响免疫细胞的数量和活性，从而改变免疫细胞和肿瘤细胞在肿瘤微环境中的相互调控作用。

免疫相关 lncRNA 通过影响肿瘤微环境作用于肿瘤的发生和进展。本文对此做一综述。

2.1 胃癌

胃癌是全球第四大最为普遍的癌症类型，也是紧随肺癌和结肠直肠癌之后的第三大致命癌症^[1, 2]。通常情况下，胃癌表现为“三高三低”病情，其中包括高发病率、高死亡率以及高转移率，同时早期诊断率、根治性切除率和5年生存率均相对较低。性别、年龄和地区是影响胃癌发病率的主要因素之一。据流行病学研究报道，胃癌在男性中排名第四，女性中排名第七，女性患胃癌的平均发病率约为男性的一半左右^[4]。以前的研究表明，胃癌在50岁以下的成年人中非常罕见，但最近的研究发现，早发性胃癌（EOGC）在年轻人群中呈上升趋势^[5]。随着人群年龄的增长，胃癌的发病率也呈上升趋势。中国快速老龄化的人口将给胃癌的医疗资源带来额外负担^[6, 7]。

2018年的数据显示，东亚地区是全球胃癌病例数量的主要集中地，超过60%的病例分布在该地区，尤其是中国、韩国和日本等国家^[8, 9]。胃癌的发病率在同一国家内不同地区间存在显著差异。例如，在我国，东部沿海地区和北方地区的胃癌发病率明显高于其他地区^[8, 10-12]。这种地理差异可能与危险因素的不同有关，例如幽门螺旋杆菌感染的差异可能是重要原因。据研究表明，感染幽门螺杆菌会显著增加胃癌的发病风险，风险增加约为未感染者的两倍^[13]。2017年的一项流行病学研究显示，幽门螺杆菌在我国人群中感染率高达55%，因此，根除幽门螺杆菌感染已被认为是降低胃癌发病率的重要措施之一^[14, 15]。此外，吸烟、不良饮食和饮食习惯、过量饮酒等因素也是胃癌的外在危险因素^[16-18]。研究胃癌的危险因素并加以干预和控制，如

以上内容仅为本文档的试下载部分，为可阅读页数的一半内容。如要下载或阅读全文，请访问：<https://d.book118.com/847100060110006046>