

通过单细胞测序和机器学习揭示 LY6E 作为肝癌关键预后标志物的潜力

宋语牧, 葛剑云, 郑允文

广东省大型动物模型生物医学重点实验室, 华南大型动物模型生物医学研究院, 中国广东省江门市五邑大学药学与食品工程学院 江门广东

【摘要】目的 肝癌, 作为一种全球性的严重健康威胁, 其发病机制错综复杂, 而传统治疗方法效果有限。生物信息学和机器学习技术的融合, 为在基因数据库中识别与疾病相关的基因提供了可能, 为开发创新治疗方法和确定新的靶点提供了潜在途径。**方法** 下载人肝癌和正常的单细胞转录组测序数据, 通过 R 程序整理分析找到在正常和疾病组差异较大的基因, 通过多种机器学习进行筛选获得核心基因, 然后验证核心基因的预后效果并检验其与免疫细胞和功能的相关性。**结果** 我们发现一批显著变化的基因。借助机器学习分析, 提炼并确认了 11 个核心基因。同时, 我们预测了核心基因的表达水平与患者的生存时间的关联揭示了核心基因与免疫细胞之间的紧密关系。**结论** 我们的研究在单细胞水平上展示了人的肝细胞癌模型, 揭示了癌变肝脏的关键特征, 包括 T 细胞 NK 细胞的增多并通过机器学习挑选核心基因并证实了 LY6E 可能是影响肝癌进展的关键基因。

【关键词】 肝细胞癌; 机器学习; 单细胞测序

【基金项目】 东京大学医学研究所国际合作研究项目资助; 国家自然科学基金 (No82070638, No82270697)

【收稿日期】 2024 年 2 月 17 日

【出刊日期】 2024 年 3 月 25 日

【DOI】 10.12208/j.ijcr.20240091

Revealing the potential of LY6E as a key prognostic marker for hepatocellular carcinoma by single-cell sequencing and machine learning

Yumu Song, Jianyun Ge, Yunwen Zheng

Guangdong Provincial Key Laboratory of Large Animal Models for Biomedicine, and South China Institute of Large Animal Models for Biomedicine, School of Pharmacy and Food Engineering, Wuyi University, Jiangmen, Guangdong

【Abstract】Objective Liver cancer, a serious global health threat, presents complex pathogenesis and limited efficacy of traditional treatments. The integration of bioinformatics and machine learning offers a pathway to identify disease-related genes in genetic databases, paving the way for developing innovative treatments and identifying new targets. **Methods** Human liver cancer and normal single-cell transcriptome sequencing data were downloaded and analyzed using R programming to identify genes with significant differences between normal and disease groups. Various machine learning methods were employed to screen for core genes, which were then validated for their prognostic significance and examined for their association with immune cells and functions. **Results** A set of significantly altered genes was identified. Through machine learning analysis, 11 core genes were refined and confirmed. Additionally, the association between the expression levels of core genes and patient survival time was predicted, revealing a close relationship between these genes and immune cells. **Conclusion** Our study presents a human liver cancer model at the single-cell level, revealing key characteristics of cancerous liver, including the increase of T cells and NK cells. Machine learning was utilized to select core genes, confirming that LY6E may be a critical gene influencing the progression of liver cancer.

【Keywords】 Hepatocellular carcinoma; Machine learning; Single cell sequencing

肝癌是一种源于肝脏的细胞的恶性肿瘤。主要包

括肝细胞癌、肝内胆管细胞癌、肝硬化后癌、肝转移瘤、小肝细胞癌和肝血管瘤等。肝细胞癌是最常见的原发性肝癌类型, 也是全球癌症相关死亡的第四大原因^[1],

已被认为是肝硬化患者死亡的主要原因^[2]。肝细胞癌的发展与慢性肝病密切相关^[3]。大多数肝细胞癌病例发生在撒哈拉以南非洲和东亚, 那里的主要危险因素是慢性乙型肝炎和黄曲霉毒素, 然而在欧美等地, 丙型肝炎和过量饮酒是肝细胞癌的主要危险因素^[4,5]。肝细胞癌的诊断与治疗在现代医学中依然面临重大挑战。尽管近年来在成像技术和分子生物学领域取得了显著进展, 肝细胞癌通常在疾病晚期才被确诊, 届时患者的治疗选择和预后可能已经受到限制。此外, 肝细胞癌的治疗复杂多样, 包括手术切除、化疗和放疗等, 但由于肿瘤的异质性及常伴有的肝脏疾病(如肝硬化), 治疗效果往往有限, 且复发率高。因此, 早期诊断和开发更有效的治疗策略是提高肝细胞癌患者生存率的关键。

近年来, 单细胞测序 (single-cell sequencing) 技术和机器学习在生物医学领域的应用也日益广泛, 单细胞测序技术可以在单细胞水平上分析基因表达并识别关键基因和信号通路, 揭示组织内细胞的异质性, 为我们理解疾病发生发展的分子机制提供了新的视角^[6,7]。而机器学习算法在识别复杂数据模式方面具有独特优势, 特别是将肝细胞癌基因表达谱和机器学习技术结合, 可以通过大量数据进行深度学习和预测, 有望为肝细胞癌的早期诊断和治疗提供新的方案^[8]。本研究通过大量的转录组数据并结合机器学习找到肝癌发展的核心基因并鉴定其与免疫细胞和免疫功能的相关性, 为开发新型治疗提供理论支持和潜在靶点。这对改善肝癌的临床治疗, 提高患者生活质量具有重要意义。

1 材料与方法

1.1 数据来源与预处理

GEO 数据库 (<https://www.ncbi.nlm.nih.gov/geo/>) 是一个国际公认的公共存储库, 储存了丰富的高通量测序数据, 成为收集 RNA 测序数据的重要平台^[9]。通过检索 GEO (Gene Expression Omnibus) 数据库, 选定 GSE162216 和 GSE124395 作为初步分析的单细胞测序数据。同时下载 TCGA (The Cancer Genome Atlas) 的肝癌数据。我们的分析中总共使用了 2 个样本, 其中包括 1 个正常样本和 1 个肝癌样本。scRNA-seq 数据基础分析是基于 R 环境 (v.4.2.2) 并使用 Seurat (v.4.3.0) 完成的^[10]。预设条件要求每个细胞检测到的基因数目在 10 到 5000 之间, 同时设定线粒体基因的比例小于 20%, 并且核糖体基因的比例小于 30%。随后, 数据经过标准化和归一化处理, 再通过方差分析筛选出前 2000 个具有高度可变特征的基因。同时, 对这些单细胞数据进行了主成分分析并通过 Harmony 方

法进行批次效应的去除^[11]。此外, 还找到了包括 GSE36376 和 GSE76427 在内的数据, 这些数据将被用于生存分析。

1.2 细胞注释和差异分析

对预处理后的数据, 用 UMAP (Uniform Manifold Approximation and Projection) 算法并选取前 10 个主成分进行整体降维分析。为了确保细胞注释的准确性, 我们利用 Find All Markers 函数来识别每个簇中的差异表达基因。同时使用 SingleR 包中的 Human RNA seq Data 数据, 以辅助降维结果的注释^[12], 最后, 结合 CellMarker 数据库^[13]和之前的研究, 最终确定了 7 类细胞。并采用 Finfmakers 函数进行差异分析。通过统计对照组和实验组的不同细胞占比, 我们识别出了比较重要的细胞群体并利用 ggplot2 包进行富集结果和比例图的绘制。

1.3 基因富集分析

为了探索正常组织与癌症组织中的基因表达差异, 我们选取了至少两倍表达差异且具有显著性 ($p < 0.05$) 的基因进行进一步分析。使用 clusterProfiler 包的 enrichGO 和 enrichKEGG 函数, 我们对癌症组高表达的基因进行了 GO (Gene Ontology) 功能和 KEGG (Kyoto Encyclopedia of Genes and Genomes) 通路的富集分析, 旨在揭示这些基因在癌症发生和发展中可能涉及的生物过程和代谢通路。

1.4 机器学习确定核心基因

以 TCGA 下载的正常和肝癌数据为训练集和验证集。从 315 个肝癌高表达基因出发, 通过 R 语言的内置函数 glm 对每个基因进行单因素逻辑回归分析, 并计算每个基因的比值比及其 99% 的置信区间且 p 值小于 0.05 的基因被认为有显著影响, 高表达基因的数量从 315 个降至 165 个。在进行单因素逻辑回归筛选后, 通过 glmnet 包利用 glmnet 函数在二项分布家族下对数据进行了 LASSO 回归分析, 采用了 10 折交叉验证方法, 通过最小化交叉验证误差来确定最优的正则化参数 λ 。最终确立 11 个关键基因。利用 mlr3verse 包对各个基因使用了包括支持向量机 (SVM)、迭代二分法 3 代 (Iterative Dichotomiser 3)、逻辑回归 (Logistic Regression)、随机森林 (Random Forest)、决策树 (Decision Tree) 和朴素贝叶斯 (Naive Bayes) 在内的一系列机器学习算法进行了预测分析, 并进行 5 折交叉验证, 重复为 10 次, 最终支持向量机 (SVM) 的预测效果最为出色。并验证了所挑选的 11 个核心基因的预测效果。

1.5 生存分析和免疫评估

在我们的研究中, 我们采用 R 语言进行 Kaplan-Meier 生存分析, 以探索特定基因在肝细胞癌中的表达差异及其对患者生存的影响。首先, 我们从公开数据库读取肝癌和对应正常组织的基因表达数据, 对基因表达水平进行预处理, 将样本根据核心基因表达量的中位数分为高表达组和低表达组。

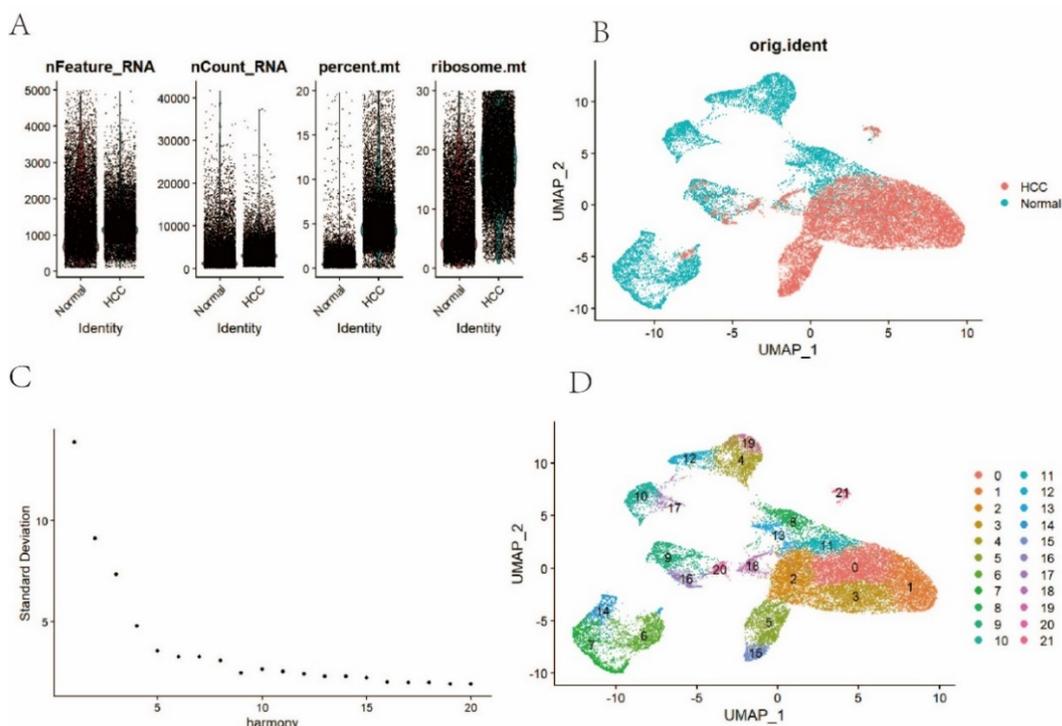
接着, 我们利用生存分析包 `survival` 进行生存函数拟合, 并使用 `survminer` 包对结果进行可视化, 生成 Kaplan-Meier 生存曲线图。通过 `GSVA` 包 `gsva` 函数执行 ssGSEA 分析验证核心基因在 Bulk RNA-Seq 数据的 ssGSEA 得分^[14], 进而评估核心基因与免疫细胞和免疫

功能的相关性。

2 结果

2.1 数据预处理

我们的对获得的单细胞测序数据进行标准的质量控制 (图 1A)。预处理后, 我们进行了和谐算法集成, 以有效集成并消除批次效应。harmony 算法后集成的 2 个样本的情况 (图 B)。其中 1: 10 主成分和基于 harmony 的批次处理用于下游分析, (图 1C)。按照 Seurat 标准步骤, 我们最终识别出 22 个簇, 然后将其嵌入到 UMAP 中 (图 1D)。最终包含 26870 个细胞 (包括 1 个正常样本和 1 个 肝细胞癌样本) 用于进一步分析。



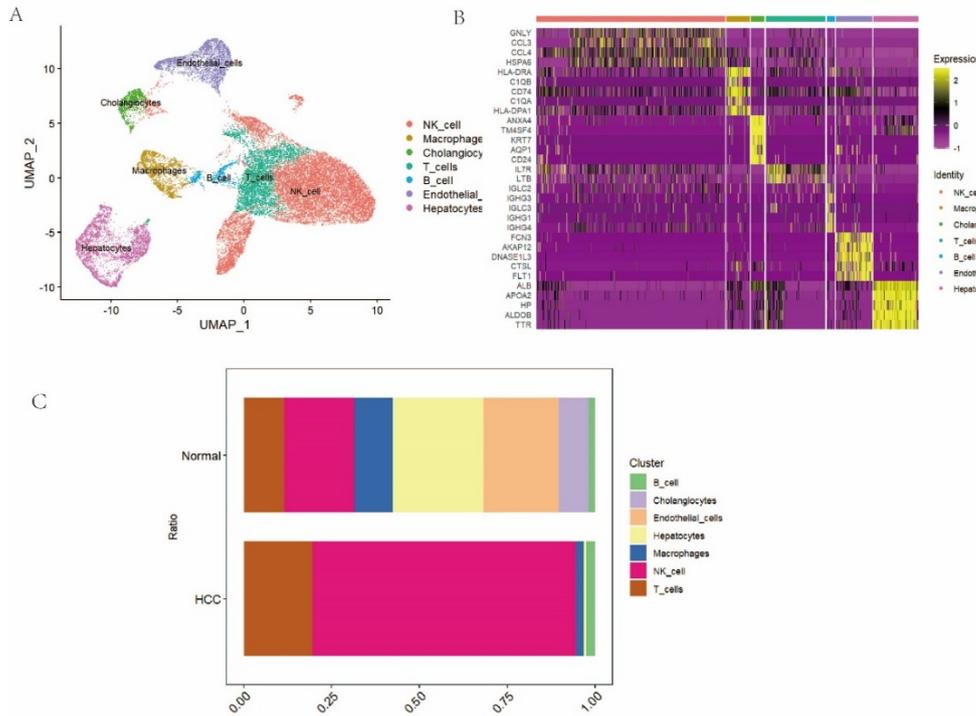
(A) 每个分析样品的特征、线粒体基因和核糖体基因的百分比; (B) UMAP 图可视化了使用 harmony 算法后每个样本的分布情况; (C) 选择数据中最能解释变异的前 15 个成分。(D) UMAP 降维聚类的分布结果

图 1 用于下游分析的 scRNA-seq 数据预处理

2.2 人肝癌的单细胞转录组分析

从 GEO 数据库获得的人类肝癌和正常的单细胞测序数据, 经过前期预处理获得 26870 个细胞, 其中 Normal 有 12164 个细胞而实验组 HCC (Hepatocellular Carcinoma) 有 14706 个细胞。我们使用 UMAP 进行数据的降维、聚类、可视化处理。共鉴定了 22 个 clusters。通过分析每个簇的标记基因, 将 32322 个细胞注释为 7 种细胞类型, 包括内皮细胞, 肝细胞, 巨噬细胞, B 细胞, T 细胞, 胆管细胞和 NK 细胞 (图 2A)。细胞

的标记基因通过热图展示 (图 2B)。肝癌组和正常组细胞比例图 (图 2C) 显示相较于正常组 NK 细胞和 T 细胞在肝癌组的比例大幅度上升。原因是 NK 细胞作为重要的免疫系统成员, 它们参与免疫监视, 能够识别和消灭早期肿瘤细胞。在肝癌组织中, NK 细胞的比例上升可能是免疫系统对肿瘤细胞增殖的反应^[15]。而肝细胞和内皮细胞等行使肝脏正常功能的细胞比例下降, 这是由于肿瘤细胞的无控制增殖或者通过分泌特定因子诱导周围正常肝细胞的凋亡导致的。



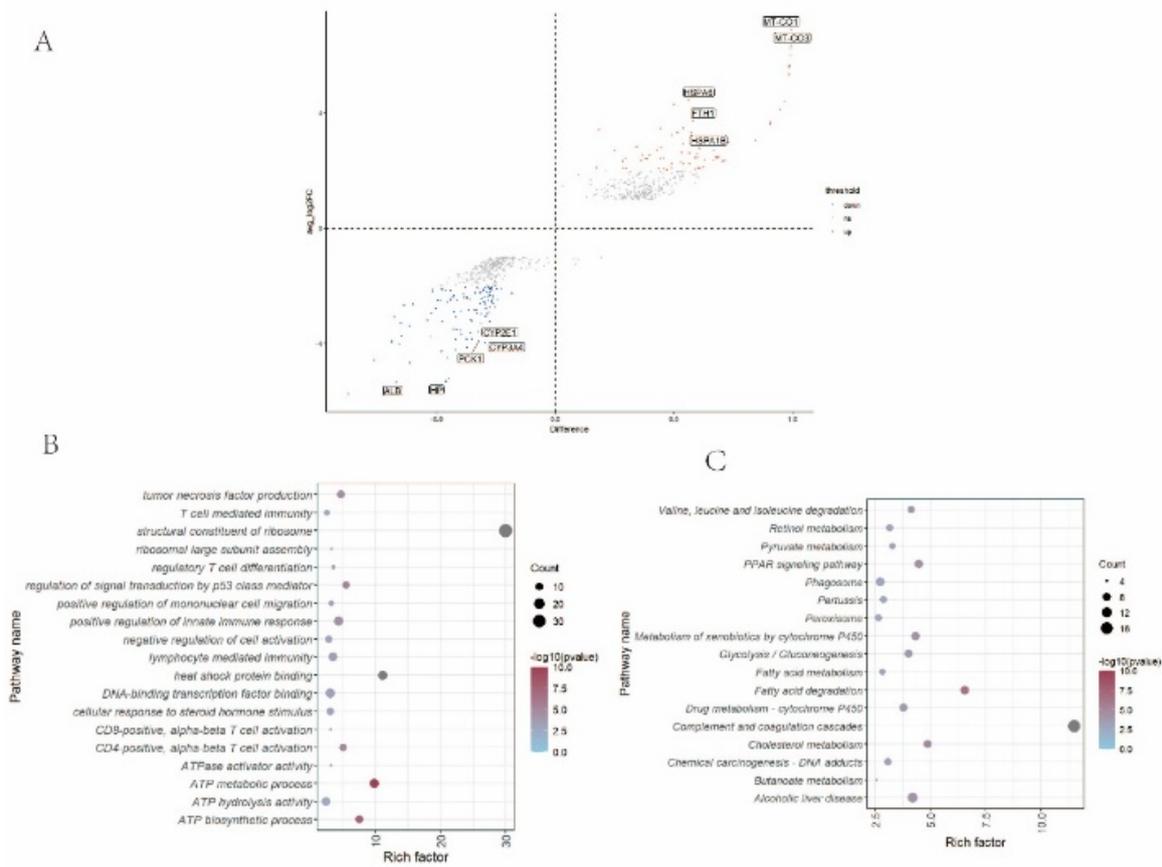
(A) 不同类型细胞的空间分布情况；(B) 不同细胞群体的标记基因展示；(C) 正常组和肝细胞癌组不同细胞分布比例

图 2 人正常肝脏和肝细胞癌的单细胞分析

2.3 候选基因与富集分析

根据以上结果，我们对正常肝脏细胞和肝细胞癌进行差异分析（差异基因有 table S1 上传）发现肝细胞癌高表达包括 MT-CO3 MT-CO2 等大量线粒体基因，这是由于肿瘤细胞特别是像肝细胞癌的代谢活跃型肿瘤，可能会调整它们的能量代谢来支持快速的生长和分裂（图 3A）。同时 HSPA6 和 FTH1 的表达量也显著高于正常组，而前者已被证实不仅在肿瘤发生和肿瘤进展中发挥作用，而且还引起非肿瘤相关疾病^[16]。而 FTH1 作为肿瘤标记物的可能也被证实^[17]。与此同时肝功能相关的基因表达量显著下调如 CYP3A4 和 CYP2E 作为肝脏中最丰富的细胞色素 P450 酶，在药物代谢中扮演重要角色^[18]，CYP3A4 的下调可能影响药物的代谢，导致治疗效果和副作用的变化。而 ALB 是肝脏合成的主要血浆蛋白质，其水平的下降通常与肝功能损害有关，可能反映了肝脏合成功能的减退。本研究通过 GO 富集分析揭示了肝癌细胞高表达基因的生物学特征（图 3B）。特别地，ATP 代谢过程和 ATP 生物合成过程的显著富集指出肝癌细胞在能量代谢方面的活跃性。这一发现与癌细胞的高能量需求和增强的增殖能力相符合。此外，由 p53 类介质调节的信号

转导的富集反映了肿瘤抑制因子 p53 在肝癌发展中的潜在作用，暗示着肿瘤细胞可能通过调节 p53 相关途径来促进生存和增殖^[19]。此外 T 细胞激活、以及肿瘤坏死因子产生的富集表明肝癌细胞可能改变宿主的免疫应答，这些变化可能涉及肝癌的免疫逃逸机制。KEGG 通路分析为我们提供了肝癌细胞代谢和信号传导网络的视角（图 3C）。其中补体和凝血途径的富集揭示了与肝癌相关的炎症反应和血栓形成可能性。这一结果指向了肝癌微环境的复杂性，强调了炎症反应在肝癌发展中的重要性。与此同时，脂肪酸降解和胆固醇代谢通路的富集强调了肝癌细胞在脂质代谢方面的变化。由于肝脏是体内主要的脂质代谢器官，这些结果提示肝癌可能伴随着肝脏正常脂质代谢功能的损害^[20]。这些通路确实表明我们筛选的差异基因与肝癌的生物学特征和病理过程紧密相关并且在肿瘤生物标志物的发现和肿瘤发展机理的研究中至关重要。肝癌细胞展现出的这些变化揭示了肝细胞癌在代谢和功能上的重编程，这不仅可能影响药物的代谢，还可能导致肝功能的损害。总体而言，这些发现不仅加深了我们对肝癌分子机制的理解，也为未来开发新的诊断方法和治疗策略提供了重要的线索。



(A) 点状图显示正常组和 肝细胞癌组中差异基因的分布情况；(B) 肝细胞癌组上调基因的 GO 富集结果；(C) 肝细胞癌组上调基因的 KEGG 富集结果

图 3 正常组和肝细胞癌组的差异分析和富集分析

2.4 机器学习确立核心基因

我们从 TCGA 数据库下载肝细胞癌的数据作为机器学习的训练集和测试集，并使用了单因素逻辑回归分析来评估各个基因对正常和肝细胞癌组别的影响。我们对每个基因进行了逻辑回归分析，如果基因的 p 值小于 0.05，我们则认为该基因对组别有显著影响，在差异基因的基础上共获得 164 个关键基因。然后我们对关键基因应用 LASSO (Least Absolute Shrinkage and Selection Operator) 回归分析并从具有统计学意义的关键基因中选择 11 个显著特征并将其相关信息收录在核心基因中 (图 4A-B)。(核心基因有 table S2 上传) 为了减小我们核心基因预测的误差，我们进行多重机器学习对比，包括支持向量机 (SVM)：优点是在高维空间表现良好，适用于复杂数据集。缺点是对大规模数据处理效率不高，且参数调优需要经验。迭代二分法 3 代 (ID3)：这是一种决策树学习算法，优点在于易于理解和实现，但容易过拟合，对缺失数据和连续特征处

理不够好。

逻辑回归 (Logistic Regression)：优点在于简单、高效，适用于线性可分的数据集。缺点是处理非线性复杂关系能力有限。随机森林 (Random Forest)：优点是准确度高，对大量数据和特征的处理能力强，对缺失数据不敏感。缺点是模型可能较大，需要较长的训练时间，且相对难以解释。决策树 (Decision Tree)：优点是模型易于理解和解释，但它们容易过拟合，特别是当树变得非常深时。朴素贝叶斯 (Naive Bayes)：优点是在处理大量特征时仍保持高效，适合文本数据。缺点是基于朴素的特征独立假设，可能在实际复杂数据上表现不佳。经过准确度，灵敏度，特异度和错误率的综合评估我们发现支持向量机的预测效果最为出色，这可能是因为 SVM 对于高维稀疏数据集的处理能力强^[21] (图 4C)。这些结果为我们在机器学习模型选择方面提供了有价值的经验，并利用 SVM 验证了我们所挑选的 11 个核心基因的预测效果 (图 4D) 并通过 ROC

(Receiver Operating Characteristic) 曲线反应预测结果。其 HSBP1 是作为热休克蛋白家族的一员被认为在癌症发生、进展、转移和抗癌治疗耐药性方面的功能。LY6E 基因代表了癌症预后不良的新型生物标志物,除了在免疫逃逸中发挥重要作用外,也是癌症进展所必需的。总之,这些发现增进了我们对肝癌复杂生物学的理解,并为未来肝癌的诊断和治疗策略的开发提供了重要的线索。针对这些癌症相关核心基因的进一步研究,可能会揭示新的治疗途径和改善患者预后的机会。

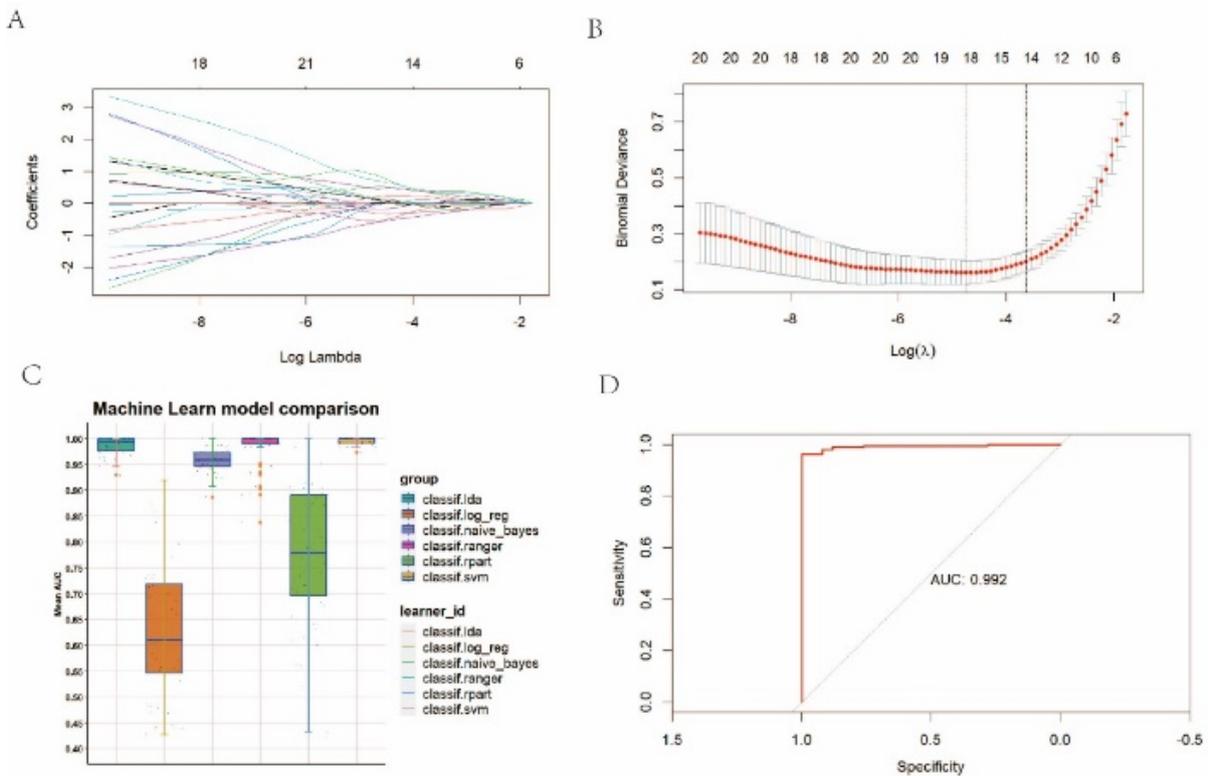
2.5 核心基因的预后效果

为了研究基因表达水平与预后之间的联系,我们对每个核心基因进行了生存关联分析,我们展示了每个核心基因在肝细胞癌患者中的表达对生存预后的影响(图 5)。图中的生存曲线清晰地区分了基因表达高低两组,揭示了它们随时间的生存率变化。其中 FOX 和 MT1X 基因的低表达组表现出更高的存活时间这说明通过抑制该基因的表达水平可以显著提高肝细胞癌患者的生存时间,同时这些基因的高表达可能促进了肿瘤的生长、侵袭或转移,从而对患者的生存产生负面影响。而尤为重要的 LY6E 基因的生存分析证明了高

表达组有更久的生存年限,这说明该基因的高表达可能与肿瘤细胞对治疗的良好响应相关联并且可能作为一种积极的生物标志物,有良好的临床预后效果。而其余基因的生存分析中高低表达组的生存曲线发生交叉对于肝细胞癌的治疗效果可能没有那么明显。总之对核心基因进行生存分析,我们发现 LY6E 作为新的治疗靶点和生物标志物的潜力,具有广泛的临床应用前景。

2.6 核心基因与免疫细胞和免疫功能相关性

在本研究中,我们通过单样本基因集富集分析(ssGSEA)来探索核心基因与不同免疫细胞的相关性并验证相关免疫功能。通过比较 LY6E 表达高低的肿瘤样本,我们发现 LY6E 高表达组与包含树突细胞与 CD8+T 细胞在内的多种免疫细胞有较强的相关性(图 6A)。此外 LY6E 也与免疫检查点和 T 细胞相关的共激活功能强相关性这说明 LY6E 在促进这些免疫功能方面有很大作用(图 6B)。这种方法允许我们深入理解 LY6E 表达对肝癌免疫微环境的影响,特别是对免疫细胞浸润和免疫反应的调控。通过详细的统计分析和数据可视化,我们能够更好地识别和解释肝癌中重要的免疫调节因素,为未来的治疗策略提供可能的靶点。



(A) 不同的颜色代表不同的基因; (B) LASSO 逻辑回归算法, 用于确立核心基因; (C) 不同机器学习结果的对比; (D) 核心基因验证的 ROC 曲线

图 4 机器学习算法确立核心基因

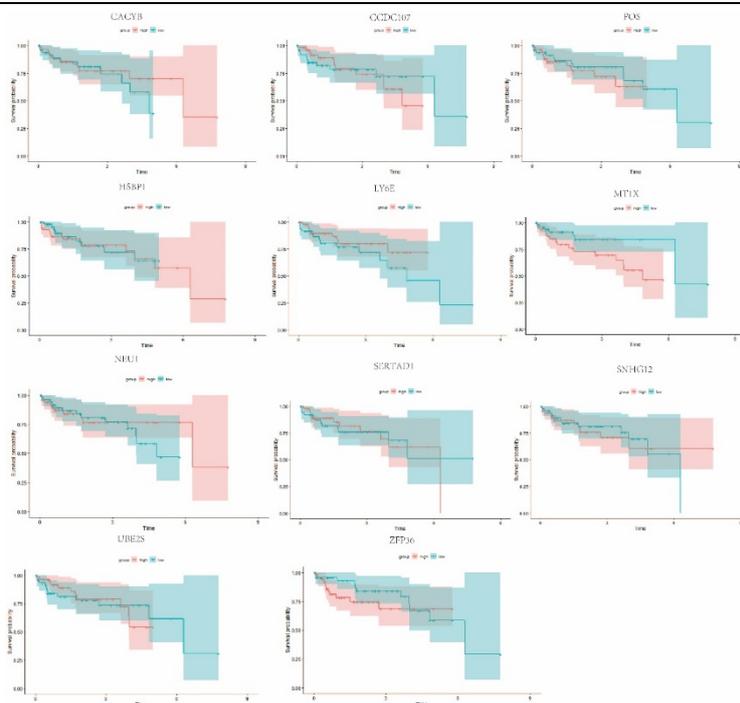
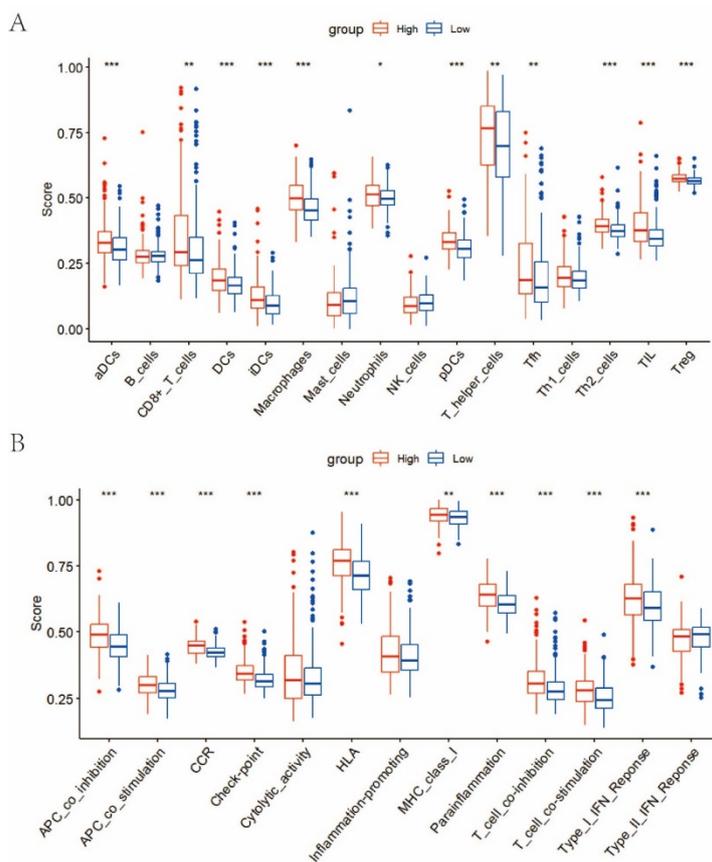


图 5 核心基因的 K-M 生存分析的预测结果



(A) LY6E 与免疫细胞的相关性; (B) LY6E 与免疫功能的相关性

图 6 LY6E 与免疫细胞和免疫功能的相关性

3 讨论

肝癌是一种发生在肝脏部位的恶性肿瘤, 随着时间的进展症状也会逐渐加重, 发展到晚期只能通过肝移植治疗。然而, 肝癌的早期诊断尚未完全清楚, 治疗手段也未完全明确, 需要深入研究。本研究利用单细胞 RNA 测序技术结合机器学习深入探索了肝细胞癌的免疫细胞和基因表达变化, 为肝细胞癌的诊断和预后提供新的见解。

我们发现, T 细胞和 NK 细胞在肝细胞癌中的比例显著增加, 这可能预示这两种免疫细胞在肝癌的发病过程中起到了关键的作用。此外, 我们通过差异分析和富集分析发现一批在正常肝脏和癌变肝脏显著变化基因, 有助于我们理解它们在疾病发展过程中的角色。在我们的研究中, 通过利用 TCGA 数据下载的肝细胞癌和正常的肝脏数据作为机器学习的训练和测试集, 我们采取了一种创新的方法来识别与人肝细胞癌相关的关键基因。使用单因素逻辑回归分析和 LASSO 回归分析筛选出具有统计显著性的核心基因并进行生存分析验证最终确立 LY6E 具有良好的标志物和治疗潜力。此外我们选择的其他核心基因尽管在 Kaplan-Meier 生存分析中与总体生存时间的直接相关性不显著, 但这并不减少对它们深入研究和实验验证的重要性, 它们可能仍对特定亚型的肝癌或在某些治疗条件下显著影响疾病进程。特别是在揭示它们在肝癌发病机制中可能的作用方面。因此, 未来的研究应包括对这些基因在不同临床背景和治疗响应下的角色的进一步探讨, 以及它们是否可以作为预测肝癌治疗效果或预后的生物标志物。这些研究可能有助于开发更个性化的治疗方案, 为肝癌患者提供更精准的治疗选择。

本研究的一个主要局限性在于对特定公共数据库, 如 TCGA 的依赖。这种依赖可能导致样本选择偏差, 限制了我们发现的普遍性和适用性。例如, 由于样本收集的地理和人口学差异, 这可能导致某些群体特有的遗传变异或表型特征未被充分考虑。此外, 数据库中的样本往往是在特定的临床条件下收集, 这可能限制了我们的分析结果对更广泛临床情景的适应性。在模型选择方面, 我们采用了单因素逻辑回归和 LASSO 回归来识别关键基因, 它们可能无法捕捉基因间的复杂相互作用和网络效应, 这对于理解复杂的生物系统至关重要。此外, 由于这些模型的参数选择和优化依赖于特定数据集, 存在过度拟合数据或忽略其他生物学重要特征的风险。此外我们的研究仅在差异基因中找核心基因, 未来还需要考察其他免疫细胞如 NK 细胞、

T 细胞等在肝癌中的作用。最后, 尽管我们的计算方法能够提供关键基因的有力证据, 但缺乏生物学实验的验证。单纯依靠计算分析得到的结果需要通过实验室条件下的功能研究来支持, 例如基因敲除或过表达实验, 以及在体内模型中的研究。这些实验不仅可以确认计算发现的基因在肝细胞癌中的作用, 还能进一步揭示它们的具体机制和潜在的治疗靶标作用。为了克服本研究在数据集依赖性和模型选择上的局限, 未来工作应致力于综合应用跨地理和人群的多元化样本集, 以增强研究的外推有效性。此外, 采用高级计算方法, 包括深度学习算法和复杂系统建模, 将有助于捕捉基因间的非线性相互作用, 从而优化模型的预测准确性并减少偏差。至关重要, 通过实验室条件下的基因功能实验和体内模型研究, 对计算分析得出的关键基因进行生物学验证, 以确立其在肝细胞癌发病机制中的确切作用。进一步地, 对免疫微环境中不同细胞类型的作用机制进行深入研究, 将有助于识别新的治疗靶标, 并促进针对肝细胞癌的免疫调节策略的开发。

总的来说, 我们的结果揭示了一系列可能与肝细胞癌发病机制相关的基因, 这些发现为深入理解肝细胞癌的病理机制提供了新的视角。也为开发针对肝细胞癌的新的治疗策略提供了可能的靶点。

参考文献

- [1] Kanwal F, Singal AG. Surveillance for Hepatocellular Carcinoma: Current Best Practice and Future Direction. [J]. *Gastroenterology* 2019; 157(1): 54-64 [PMID: 30986389 PMCID: Q1].
- [2] Hepatocellular carcinoma. [J]. *Nat Rev Dis Primers* 2021; 7(1): 7 [PMID: 33479233 PMCID: Q1].
- [3] Morgan RL, Baack B, Smith BD, et al. Eradication of hepatitis C virus infection and the development of hepatocellular carcinoma: a meta-analysis of observational studies. [J]. *Ann Intern Med* 2013; 158(5 Pt 1): 329-337 [PMID: 23460056 PMCID: Q1].
- [4] Chen C-J, Yang H-I, Su J, et al. Risk of hepatocellular carcinoma across a biological gradient of serum hepatitis B virus DNA level. [J]. *JAMA* 2006; 295(1): 65-73 [PMID: 16391218 PMCID: Q1].
- [5] Chen J-D, Yang H-I, Iloeje UH, et al. Carriers of inactive hepatitis B virus are still at risk for hepatocellular carcinoma and liver-related death. [J]. *Gastroenterology* 2010; 138(5): 1747-1754 [PMID: 20114048 PMCID: Q1].

- [6] Potter SS. Single-cell RNA sequencing for the study of development, physiology and disease. [J].*Nature reviews Nephrology* 2018; 14(8): 479-492 [PMID: 29789704 PMID: PMC6070143].
- [7] Single-cell transcriptomics of 20 mouse organs creates a Tabula Muris. [J].*Nature* 2018; 562(7727): 367-372 [PMID: 30283141 PMID: PMC6642641].
- [8] Azuaje F. Artificial intelligence for precision oncology: beyond patient stratification. [J].*NPJ precision oncology* 2019; 3: 6 [PMID: 30820462 PMID: PMC6389974].
- [9] Barrett T, Wilhite SE, Ledoux P, et al. NCBI GEO: archive for functional genomics data sets--update. [J].*Nucleic acids research* 2013; 41(Database issue): D991-995 [PMID: 23193258 PMID: PMC3531084].
- [10] Stuart T, Butler A, Hoffman P, et al. Comprehensive Integration of Single-Cell Data. [J].*Cell* 2019; 177(7): 1888-1902.e1821 [PMID: 31178118 PMID: PMC6687398].
- [11] Korsunsky I, Millard N, Fan J, et al. Fast, sensitive and accurate integration of single-cell data with Harmony. [J].*Nature methods* 2019; 16(12): 1289-1296 [PMID: 31740819 PMID: PMC6884693].
- [12] Aran D, Looney AP, Liu L, et al. Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. [J].*Nature immunology* 2019; 20(2): 163-172 [PMID: 30643263 PMID: PMC6340744].
- [13] Zhang X, Lan Y, Xu J, et al. CellMarker: a manually curated resource of cell markers in human and mouse. [J].*Nucleic acids research* 2019; 47(D1): D721-d728 [PMID: 30289549 PMID: PMC6323899].
- [14] Hänzelmann S, Castelo R, Guinney J. GSVA: gene set variation analysis for microarray and RNA-seq data. [J].*BMC bioinformatics* 2013; 14: 7 [PMID: 23323831 PMID: PMC3618321].
- [15] Purdy AK, Campbell KS. Natural killer cells and cancer: regulation by the killer cell Ig-like receptors (KIR). [J].*Cancer Biol Ther* 2009; 8(23): 2211-2220 [PMID: 19923897 PMID: PMC2825280].
- [16] Song B, Shen S, Fu S, et al. HSPA6 and its role in cancers and other diseases. [J].*Mol Biol Rep* 2022; 49(11): 10565-10577 [PMID: 35666422].
- [17] Huang H, Qiu Y, Huang G, et al. Value of Ferritin Heavy Chain (FTH1) Expression in Diagnosis and Prognosis of Renal Cell Carcinoma. [J].*Med Sci Monit* 2019; 25: 3700-3715 [PMID: 31104064 PMID: PMC6537665].
- [18] Waring RH. Cytochrome P450: genotype to phenotype. [J].*Xenobiotica* 2020; 50(1): 9-18 [PMID: 31411087].
- [19] Mantovani F, Collavin L, Del Sal G. Mutant p53 as a guardian of the cancer cell. [J].*Cell Death Differ* 2019; 26(2): 199-212 [PMID: 30538286 PMID: PMC6329812].
- [20] Pope ED, 3rd, Kimbrough EO, et al. Aberrant lipid metabolism as a therapeutic target in liver cancer. [J].*Expert Opin Ther Targets* 2019; 23(6): 473-483 [PMID: 31076001 PMID: PMC6594827].
- [21] Zhou S. Sparse SVM for Sufficient Data Reduction. [J].*IEEE transactions on pattern analysis and machine intelligence* 2022; 44(9): 5560-5571 [PMID: 33891547].

版权声明: ©2024 作者与开放获取期刊研究中心 (OAJRC) 所有。本文章按照知识共享署名许可条款发表。
<http://creativecommons.org/licenses/by/4.0/>



OPEN ACCESS